# A new automated method for improved flood defense representation in large-scale hydraulic models

**Oliver E. J. Wing[1,2], Paul D. Bates[1,2], Jeffrey C. Neal[1,2], Christopher C. Sampson[2], Andrew M. Smith[2], Niall Quinn[2], Iuliia Shustikova[3], Alessio Domeneghetti[3], Daniel W. Gilles[4], Radoslaw Goska[4], and Witold F. Krajewski[4]**

[1]School of Geographical Sciences, University of Bristol, Bristol, United Kingdom;

[2]Fathom, Bristol, United Kingdom;

[3]DICAM, University of Bologna, Bologna, Italy;

[4]IIHR–Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa, United States.

Corresponding author: Oliver E. J. Wing (oliver.wing@bristol.ac.uk)

**Key Points:**

- Flood defense representation is presently poor in large-scale flood models, impairing their ability to map flood hazard accurately.
- A new method is presented which automatically identifies hydraulic structures in terrain data and accurately preserves their elevations.
- Hydraulic simulations where lack of defense data is the dominant error show significant improvements in skill when incorporating this method.

## Abstract

The execution of hydraulic models at large spatial scales has yielded a step-change in our understanding of flood risk. Yet, their necessary simplification through the use of coarsened terrain data results in an artificially smooth Digital Elevation Model (DEM) with diminished representation of flood defense structures. Current approaches to dealing with this, if anything is done at all, involve either employing incomplete inventories of flood defense information or making largely unsubstantiated assumptions about defense locations and standards based on socio-economic data. Here, we introduce a novel solution for application at-scale. The geomorphometric characteristics of defense structures are sampled and these are fed into a probabilistic algorithm to identify hydraulically relevant features in the source DEM. The elevation of these features is then preserved during the grid coarsening process. The method was shown to compare favorably to surveyed US levee crest heights. When incorporated into a continental-scale hydrodynamic model based on LISFLOOD-FP and compared to local flood models in Iowa (US), median correspondence was 69% for high frequency floods and 80% for low frequency floods, approaching the error inherent in quantifying extreme flows. However, improvements versus a model with no defenses were muted and risk-based deviations between the local and continental models were large. When simulating an event on the Po River (Italy), built and tested with higher quality data, the method outperformed both undefended and even engineering-grade models. As such, particularly when employed alongside model components of commensurate quality, the method here generates improved-accuracy simulations of flood inundation.



## Plain Language Summary

Traditional flood risk assessments are carried out using computer models built with local data, but their spatial coverage is impaired by how expensive and time-consuming they are to produce. Recent advances in data availability, understanding of necessary physical process representation and computational capacity have enabled hydraulic models of the entire globe to be built in an automated fashion at a fraction of the financial and human cost. However, their accuracy can be significantly impaired by a lack of information on flood defenses. As the model is built, elevation data is coarsened to reduce the number of calculations required to simulate flooding over such wide areas. This results in flood defense structures being smoothed out of the terrain information used in the model. Publicly available defense inventories are of insufficient coverage to ameliorate this issue. In this paper, a method is presented which automatically detects levee-like features in high-resolution elevation data and accurately represents their heights during this necessary coarsening process. Simulating flood inundation over this “defended” topography results in high correspondence between local models and observations for test cases in the US and Italy, with improvements particularly felt where a lack of defense information is the dominant source of error.

## 1 Introduction

The last decade has seen a revolution in the field of flood inundation modelling. Historically, hydraulicians have focused on custom-building local models of individual river reaches, but the dual effect of enhanced computational capacity and the advent of "big data" have expanded the size of model domains considered to entire regions, continents and even the globe (Dottori et al., 2016; Sampson et al., 2015; Wilson et al., 2007; Wing et al., 2017; Winsemius et al., 2013; Yamazaki et al., 2011). However, a significant performance gap exists between small- and large-scale models. Event-replicating local models can closely match the real-world observations they attempt to emulate (e.g. Bates et al., 2006; Mignot et al., 2006; Neal et al., 2009), while global-scale models deviate significantly from flood extent observations, local models and even each other (e.g. Dottori et al., 2016; Sampson et al., 2015; Trigg et al., 2016; Ward et al., 2017). Regional- and continental-scale models – with substantial, but not worldwide, model domains usually affording them more localized, higher quality data – outperform global models (e.g. Fleischmann et al., 2019; Wilson et al., 2007; Wing et al., 2017), but the "spatial scale performance gap" is still evident.

Attempts to achieve performance parity between models of all spatial scale are well underway in the field. These generally involve emulating the features of local-scale modelling strategies that enable their accurate simulation of flood inundation. This typically requires manual interventions in the model-building process by a skilled practitioner and the collection of accurate, local information: neither of which are possible when operating at large spatial scales. Viable alternatives to local-scale approaches that can be employed in a larger scale model therefore must be available at regional–global scales and incorporated into an automated model building process. Much of the research into these alternatives is either in or approaching a state

of maturity: related to the processing of elevation data, improvements to computational hydraulics, representation of rivers, characterization of the extreme flow inputs, and model validation. While calls for industry, governments and NGOs to support the development of a publicly-available, global Digital Elevation Model (DEM) built from laser altimetry (lidar) data are ongoing (Sampson et al., 2016; Schumann, 2014; Winsemius et al., 2019), corrections to errors in spaceborne radar-based DEMs have made such data fit for purpose in flood modelling at large scales (e.g. Archer et al., 2018; Yamazaki et al., 2017). The development of simplified hydraulic codes, which ensure physical process representation is commensurate with data availability and computational burden (Hunter et al., 2007), have enabled the rapid spatial "scale-up" seen in the past decade (e.g. Bates et al., 2010; Lamb et al., 2009; Neal et al., 2018; Sanders & Schubert, 2019). The delineation of channel networks from global DEMs (e.g. Allen & Pavelsky, 2018; Lehner et al., 2008) and approximations regarding their geometry (e.g. Frasson et al., 2019) ensure the necessary channel representation (Neal et al., 2012) in hydraulic models amidst data scarcity at large spatial scales. Characterizing extreme flows, given the paucity of river gauges in time and space, is commonly performed by cascading climate reanalysis data through a hydrological model (e.g. Dottori et al., 2016; Pappenberger et al., 2012) or by forging statistical relationships between catchment characteristics and gauge data and transferring these to ungauged regions (e.g. Smith et al., 2015). Finally, the validation of these model structures, inhibited by the scarcity of benchmark data at a commensurate spatial scale, is becoming more common place (e.g. Dottori et al., 2016; Ward et al., 2017; Wing et al., 2017). In contrast, a critical component of large-scale hydraulic models that has received relatively little attention is the representation of flood defense structures within them.

Information on the location of flood defense structures on the floodplain (specifically levees, embankments, dykes, flood walls: these terms are used interchangeably hereafter) at large spatial scales is very scarce; metadata regarding their defense standards or crest heights are scarcer still. The horizontal resolution of the elevation data employed in large-scale flood models is typically too low to fully capture the effect of such structures. Even if the source elevation data is higher resolution (e.g. airborne lidar surveys which are usually available on 1–5 m grids), the resampling process required to reduce the resolution to a scale tractable for large-scale simulations imparts a smoothing effect that reduces the crest height of levees, often to the point of their disappearance. In light of this, the explicit effect of flood defenses is virtually ignored by large-scale flood hazard modelers. If accounted for at all, many modelers treat flood defenses as an afterthought by simulating the undefended case only and assuming floods smaller than a specified magnitude are unimportant when calculating risk (Alfieri et al., 2017; Dottori et al., 2018; Feyen et al., 2012; Quinn et al., 2019; Winsemius et al., 2013). FLOPROS (Scussolini et al., 2016) is a commonly used dataset which provides design, policy and modelled flood defense standards globally at the sub-country level to inform the defense threshold above which floods cause damage. It is, however, not justified to assume defense standards remain constant across such vast areas, nor that every river within a region is afforded some standard of protection. Furthermore, representing the hydraulic effect of levees during flood events is shown to have a significant effect on peak flows, both upstream (Heine & Pinter, 2012) and downstream (Castellarin et al., 2011a; Di Baldassarre et al., 2010) of the levee, and the assumption that an undefended simulation behaves in the same way as in a simulation where a levee has been overtopped is not valid (e.g. Ciullo et al., 2019; Masoero et al., 2013).

126 The continental-scale model presented in Wing et al. (2017) incorporated the US Army
127 Corps of Engineers (USACE) National Levee Database (NLD) into its structure, though a report
128 by the American Society of Civil Engineers (2017) indicated that only 48,000 of an estimated
129 160,000 km (30%) of the nation's levees are contained in the USACE database. Crest heights
130 and defense standards are only sporadically specified also, meaning assumptions must be made
131 to fill these gaps: only knowing the location of a levee is not enough information for its explicit
132 inclusion in a dynamic simulation. "Burning" crest heights into the DEM is not straightforward
133 though, owing to the specification of how the levee metadata interacts with the terrain data.
134 Datum conflicts, geolocation errors and the effect of residual levee artefacts in the baseline
135 topographic data can result in an improper representation of defense information, even in the
136 relatively few locations where such data are available. If defense standards are specified (e.g. as
137 an annual exceedance probability (AEP) flood that the structure defends against), to incorporate
138 these standards in a hydrodynamic simulation a crest elevation will still be required. For the
139 model structure in Wing et al. (2017), this can be derived by linking the AEP defense standard to
140 the bank height of the river channel (Neal et al., 2012) via the flood frequency analysis which
141 determines particular AEP discharges (Smith et al., 2015). Although computationally efficient,
142 this process ignores the defensive potential of lateral floodwater storage between the channel and
143 the levee; which in many cases can be substantial (Hooijer et al., 2004). The solution to this
144 employed in Wing et al. (2017) is to run a preliminary "defense height estimation" simulation for
145 certain AEP design floods, tracking the height of water at the levee location and elevating the
146 crests to this for the primary model run. The upshot of these issues – most overwhelmingly, the
147 30% levee capture rate – is impaired performance in the Wing et al. (2017) model. For this
148 model, false alarms were reportedly higher in urban, rather than rural, areas, resulting in lower

model skill in the very areas where accurate risk calculations are most urgently required, with the likely culprit being inadequate defense representation leading to inundation in areas that are protected in reality.

Amidst incomplete defense inventories, some modelers make assumptions about where levees are likely to be situated based on socio-economic or land-use data; the general idea being that wealthier, built-up areas are afforded a higher standard of protection compared with the inverse case (Feyen et al., 2012; Quinn et al., 2019; Sampson et al., 2015). Using the US as an example, we demonstrate that such assumptions are generally invalid. The USACE NLD contains a wealth of levee data for the contiguous US, including ~200,000 $km^2$ of land specified as being defended from flooding. Sampling the characteristics of these lands in terms of degree of urbanity (from the National Land Cover Database), wealth (Median Household Income from the US Census Bureau) and government spending (USACE spending under the American Recovery and Reinvestment Act of 2009) offers insights into whether a crude predictive defense model based on nationwide socio-economic characteristics is valid. The left-hand side of Fig. 1 illustrates the correlation between the defense standard of the levee protecting an area of land and its characteristics in terms of urbanity, wealth and spending respectively. It is clear from Fig 1a that, of areas that are offered some degree of structural flood protection, urban areas are not protected to a higher standard than less developed areas. Similarly, defended wealthier neighborhoods are not offered greater protection than defended poorer ones (Fig. 1b). Indeed, very high return period levees (i.e. those that defend against the 1 in 1000 year flood) appear to defend poorer areas. Further, higher USACE spending since 2009 does not appear to be correlated with defense standard: if anything, the opposite case is true (Fig. 1c). The histograms on the right-hand side of Fig. 1 show the distribution of these characteristics within defended

areas and in the remaining US land surface outside of them. The distribution of urbanity (Fig. 1d), wealth (Fig. 1e) and spending (Fig. 1f) are virtually indistinguishable between defended and undefended lands, indicating such variables would be inappropriate for use in a statistical model to predict the location and standard of defenses.

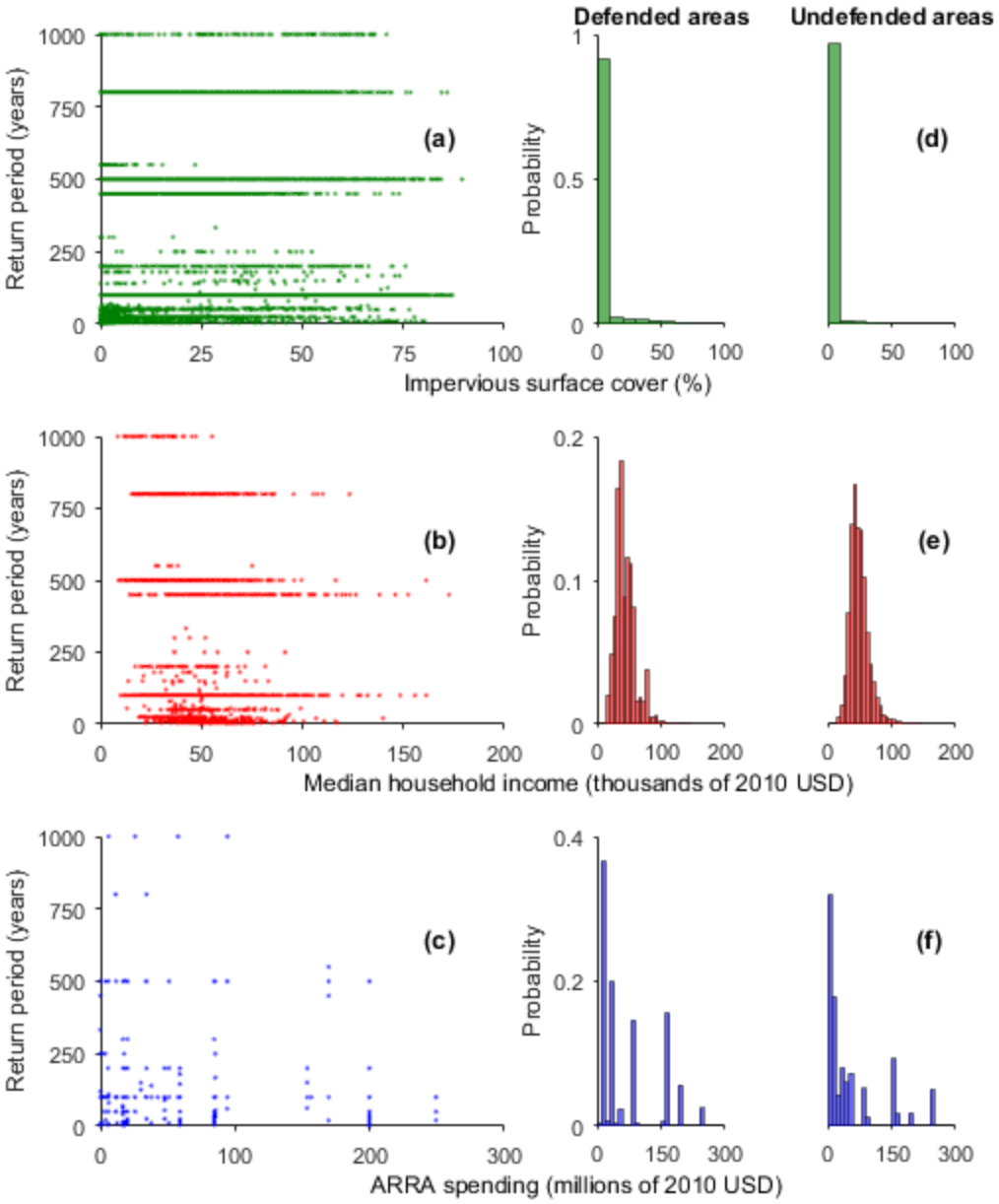



**Figure 1.** The relationship between USACE NLD defense standard and **(a)** urbanity, **(b)** wealth, and **(c)** spending, as well as the distribution of **(d)** urbanity, **(e)** wealth and **(f)** spending within and outside USACE NLD defended areas.

In this paper, we propose a novel advance in defense representation capable of integration in automated large-scale model building frameworks. Using the Fathom-US model (Wing et al., 2017) as a springboard from which to make these advances, we take the source data of the DEM employed and apply an algorithm to preserve levee crest heights during the DEM resampling process. This DEM coarsening is a necessary step in ensuring the model grid scale is computationally tractable given the vast spatial coverage of the model domain. To prevent loss of information during this process, the method automatically ensures flood defense structures remain represented in the terrain data employed. The supposition that levees can be detected based on their topographic signature is not a new one: flood defenses have been extracted from DEMs in a number of studies (Bailly et al., 2008; Casas et al., 2012; Cazorzi et al., 2013; Choung, 2014; Krüger, 2010; Krüger & Meinel, 2008; Sofia et al., 2014; Steinfeld et al., 2013), while feature extraction from DEMs more generally is a relatively well-established field (e.g. Lashermes et al., 2007; Passalacqua et al., 2010, 2012; Sofia et al., 2011; Tarolli et al., 2012). The characteristics of levees with relation to elevation and its derivatives (Evans, 1979, 1980; Wood, 1996) are distinct, meaning they can be described by geomorphometric parameters (after Sofia et al., 2014). All previous studies are broadly similar in their characterization of levees with such parameters. To actually isolate levee features though, qualitative descriptions of these elevation-derived parameters (e.g. “linear features of high elevation bounded by two opposing steep slopes”) must be translated to quantitative definitions of the parameter thresholds at which certain DEM pixels are considered relevant. It is here that previously cited studies fall short: their choice of deterministic geomorphometric parameter thresholds relevant only to the geographical domain and grid resolution of each study-specific DEM limits the applicability of their findings more widely (e.g. Choung, 2014; Krüger, 2010; Krüger & Meinel, 2008; Sofia et al., 2014).

Given these studies: (i) employed lidar-derived DEMs with resolutions of the order $10^0$ m, which are currently unavailable at large spatial scales; (ii) were of small, isolated test cases; (iii) offered little indication regarding the computational feasibility of applying such methods at larger scales; and (iv) were not extended to analyze fitness-for-purpose in a hydraulic modelling context, such methods leave crucial research questions unanswered. Other novel approaches, including those with a post-processing step to fill-in line breaks in DEM-based approaches (e.g. Choung, 2014), those related to processing image spectra, texture and shape (e.g. Steinfeld et al., 2013), and tracking observed flood edges alongside river gauge data to infer levee presence (e.g. Wood et al., 2018), are heavily impaired by data availability and only delineate the location of levees. For use in flood modelling, a view of crest elevation or defense standard is required as well. As such, the method presented here adheres to a DEM-based approach since: (i) elevation data is available at large spatial scales, albeit at a coarser resolution than in previous levee detection studies; and (ii) detection this way implicitly extracts a crest elevation in the form of peaks in the DEM, avoiding the need for further processing or assumptions to implement the results in a hydraulic model.

We initially test the accuracy of this approach by comparing a new "defended" DEM to surveyed levee crest elevations in the state of California, calculating whether the method captures crest heights more accurately than standard DEM resampling approaches. We then incorporate this approach into the Fathom-US hydrodynamic modelling framework and compare hydraulic simulations to a broad-scale amalgamation of high-quality local models in the state of Iowa, charting the degree to which local- and large-scale models converge in their simulation of design AEP flood events with improved flood defense representation. We finally then validate

227 the application of the method in a high-quality real event simulation of the well-defended Po
228 River floodplain in northern Italy.

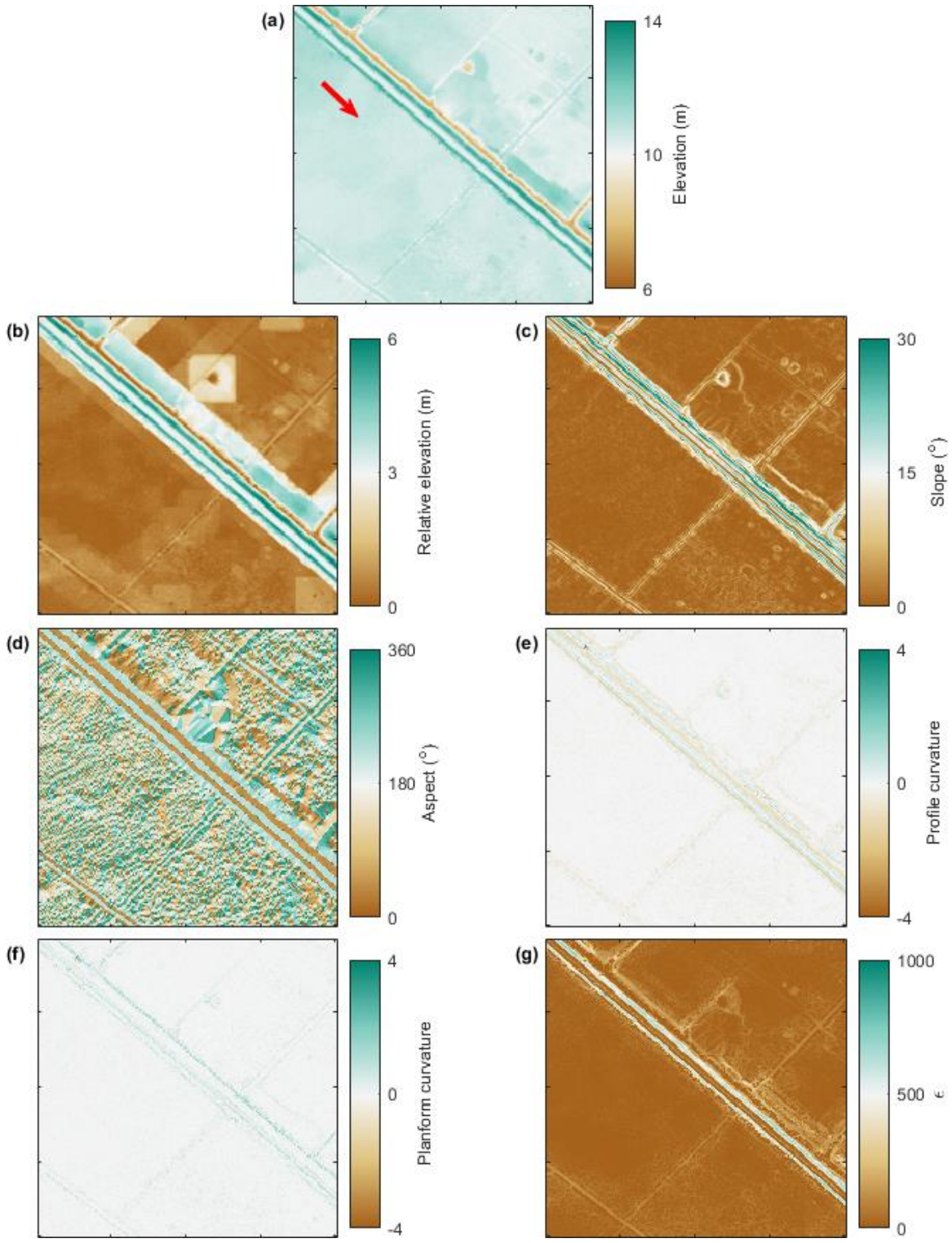
229

230 **2 Data and Methods**

231 For this study, the test area is the entire contiguous United States (CONUS), owing to the
232 availability of total-coverage elevation data from the US Geological Survey (USGS) National
233 Elevation Dataset (NED) at ⅓ arc sec (~10 m) and partial coverage at ⅑ arc sec (~3 m). The
234 NED is based on lidar data for 39% of the CONUS, containing 67% of its population. The
235 hydraulic model structure that this levee extraction method will be integrated within is Fathom-
236 US (Wing et al., 2017), which uses the NED at 1 arc sec (~30 m) as its DEM. Halving the grid
237 resolution increases computation time by an order of magnitude (Savage et al., 2016), making
238 running Fathom-US on the higher-resolution NED variants computationally intractable at
239 national scales. The proposed method intends to preserve hydraulically important information at
240 these higher resolutions for use within the 1 arc sec model structure. Fathom-US hydraulics are
241 based on a version of the LISFLOOD-FP numerical scheme: a 2D simplification of the shallow
242 water equations which approximates local inertia (Bates et al., 2010; de Almeida & Bates, 2013).
243 AEP flows, based on the regional flood frequency analysis of Smith et al. (2015) using USGS
244 river gauges, are routed through 1D subgrid channels (Neal et al., 2012), which are derived from
245 HydroSHEDS hydrography data (Lehner & Grill, 2013), and across NED-derived floodplains.
246 Fathom-US also has a pluvial model component, where the direct effect of intense rainfall onto
247 the land surface is simulated. Boundary conditions here are informed by NOAA Intensity-
248 Duration-Frequency relationships. Further details are available in Wing et al. (2017) and
249 Sampson et al. (2015).

Five geomorphometric parameters relevant for the identification of levees were sampled from the ⅓ and ⅑ arc sec NED variants: relative elevation, slope, aspect, profile curvature and planform curvature. The crucial component of this method is the automated sampling of the parameter thresholds from known levee locations, essentially "training" the extraction algorithm against ground truths derived from the USACE National Levee Database. This is another reason why the CONUS is an ideal testbed, since it contains wide-area levee information in the form of the NLD. While not containing every levee in the US, the database contains 48,000 km of flood defense locations. Geomorphometric parameters inherent in the NED at NLD locations form the parameter thresholds required to identify levees elsewhere in the NED which are not in the NLD. The algorithm inevitably captures any features within the NED that exhibit the geomorphometric characteristics of levees in the NLD: meaning other 'informal' features, which may still be of hydraulic relevance, are also captured. In recognition that no single set of parameter thresholds will adequately capture levee or levee-like features in the elevation data, a random sample (n = 1000, in this case) of the geomorphometric characteristics evident in the NLD drives the extraction algorithm 1000 times. This number is likely in excess of what is required to ensure beyond-adequate sampling of the parameter space. This generates a pseudo-probabilistic surface, assigning each DEM pixel an "extraction rate" (ε) out of 1000. Specifically, ε is the number of times a pixel exceeds all five thresholds of a given parameter set. Where ε is greater than a given threshold (between 0 and 1000), its elevation value will become the corresponding coarser 1 arc sec DEM pixel value. Where multiple higher-resolution pixels with ε above the threshold fall within a 1 arc sec cell, the maximum elevation value is taken. In cases where ε is below the threshold, some central tendency of elevation values of the higher-resolution pixels which constitute a single 1 arc sec pixel will be used. To this end we use bilinear resampling, an

aggregation process shown to produce the most accurate and physically realistic results in an analysis by Fewtrell et al. (2008). The five geomorphometric parameters, their application in the detection algorithm and their 1000 NLD-derived thresholds are detailed as follows (see Figs. 2 and 3):




277

278 **Figure 2.** Examples of a levee in a DEM and its processing in the extraction algorithm

279 for a 0.6 x 0.6 km area of Brazoria County, Texas (ticks are spaced 150 m apart). **(a)** an

area of raw ⅑ arc sec USGS NED DEM, with two levees bounding a river channel (in dark green) with flow direction indicated by the red arrow; **(b)** relative elevation, showing the levees' clear height differential above their neighborhood minima; **(c)** slope, where the regions of high relative elevation are bounded by steep slopes on either side; **(d)** aspect, illustrating that these slopes broadly face opposite directions; **(e)** profile curvature, with the levee crests exhibiting surface convexity across the direction of maximum slope; **(f)** planform curvature, showing negligible change in slope along the levee crests; **(g)** output of the extraction algorithm when driven with geomorphometric parameter thresholds on the right-hand side of Fig. 3, where the levees have much higher ε values than other terrain features.

(i) Relative elevation

Relative elevation ($z_r$) is first computed to refine the number of pixels upon which later parameters are computed (e.g. adjacent slopes are only calculated for elevated pixels). It is defined as the difference between elevation ($z$) at point ($x$,$y$) and the neighborhood minimum. Neighborhood in this context is defined as a 100 x 100 m square with point ($x$,$y$) at its center. This kernel size is selected so that, should point ($x$,$y$) be a levee crest, a representation of the true ground surface is captured for an accurate $z_r$ calculation (see Eq. 1).

$$z_r = z_{xy} - z_{min} \quad (1)$$

Our expectation is that levee crests are elevated with respect to their neighborhood minimum, resulting in a positive $z_r$. A visual example is shown in Fig. 2b. Figs. 3a–b show the 1000 thresholds at ⅓ and ⅑ arc sec respectively, derived from known levee locations. Although

both provide quantities in line with the above expectation, their distributions are different. At ⅓ arc sec, roughly 3x as many $z_r$ thresholds are < 1 m than at ⅑ arc sec. This seems to suggest that, even at ~10 m resolution, the representation of levee crests in the DEM has been diminished.

(ii) Slope

Slope ($S$) is a first-order differential of elevation: defined here as the maximum rate of change of elevation between neighboring pixels in $x$ and $y$ directions in degrees (Eq. 2). Opposing slopes are permitted to be a maximum of 20 m apart (i.e. levee crests can be 20 m wide). At opposite sides of the levee crest, we would expect to see DEM pixels with a positive $S$ (see Fig. 2c).

$$S = \tan^{-1}\left(\sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}\right) \quad (2)$$

From Figs. 3c–d, this is shown to be the case in the $S$ thresholds selected. There is a similar and expected trend as with $z_r$ in that the coarser DEM produces flatter slope thresholds than the finer one.

(iii) Aspect difference

Aspect ($A$) is another first-order differential of elevation, indicating the direction of maximum slope ($S$). This is defined by Eq. 3 in degrees clockwise from north. In the case of levees, we would expect the opposing slopes to be facing polar opposite directions (see Fig. 2d).

$$A = \tan^{-1}\left(\frac{\left(\frac{dz}{dy}\right)}{\left(\frac{dz}{dx}\right)}\right) \quad (3)$$

The thresholds are defined in terms of the differences between the aspects of opposing slopes. As in Eq. 4, where (*x*,*y*) is a potential levee crest and *m* and *n* are the distance of the opposing slopes from this crest in *x* and *y* directions respectively:

$$A_{diff} = \left| A_{(x+m,y+n)} - A_{(x-m,y-n)} \right| \quad (4)$$

The geomorphometric parameter thresholds in Fig. 3e–f are the deviation of $A_{diff}$ from 180°, permitting some tolerance in the definition of "facing the opposite direction". In most cases, this tolerance is close to 0°. The slight rise in cases where the tolerance is close to 180° (indicating $A_{diff} \approx 0$) are probably due to NLD geolocation errors where the crest is placed on the slope. In such circumstances, the steep slopes either side of this misidentified crest would be facing the same direction.

(iv) Profile curvature

Curvature is a second-order differential of elevation. Profile ($prof_c$) curvature is the rate of change of slope in the maximum downslope (*S*) direction, defined in Eq. 5:

$$prof_c = \frac{dS}{dxy} \quad (5)$$

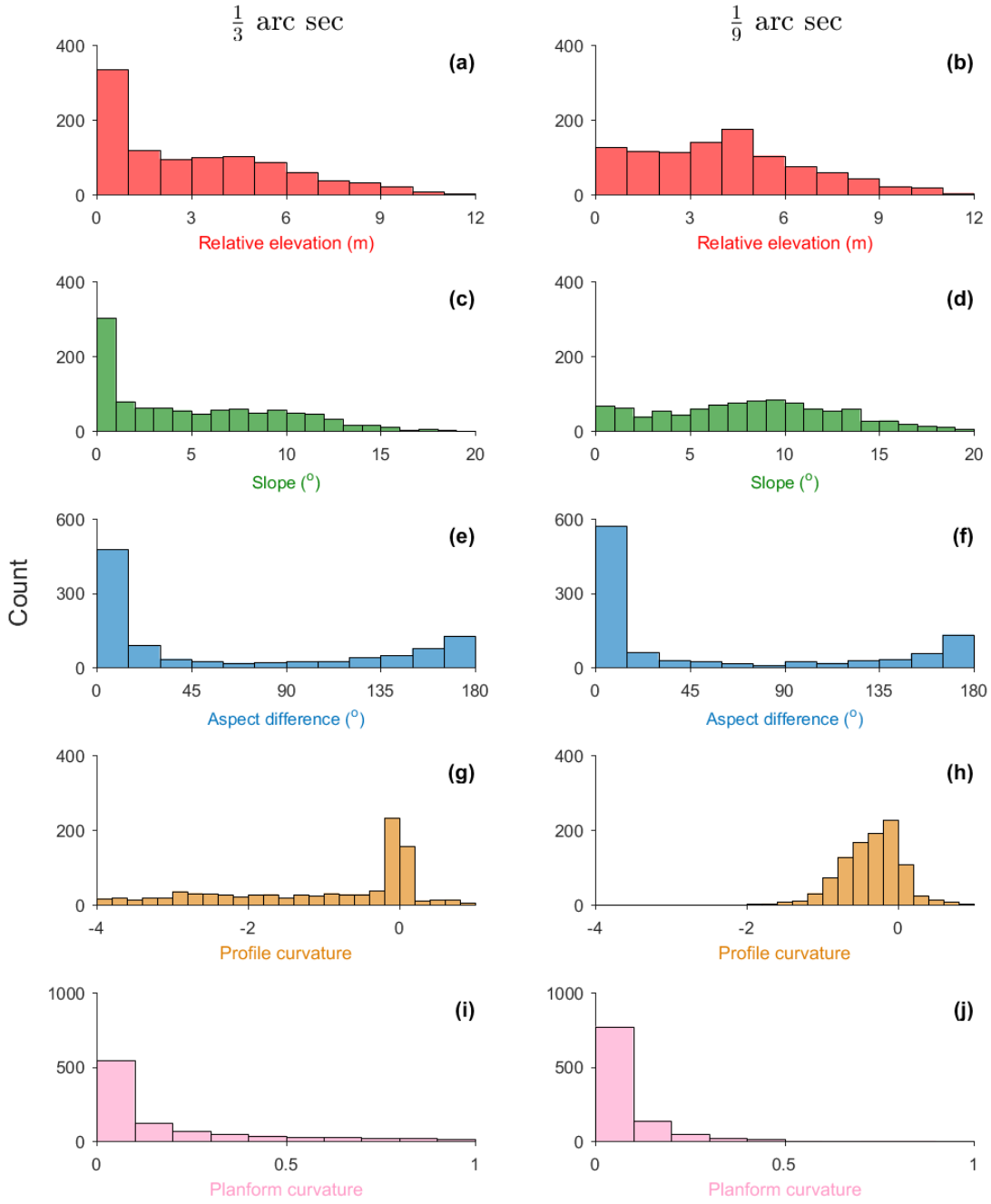
Levees are characterized by profile convexity (Fig. 2e), in this case being represented as negative profile curvature. In almost all cases, the algorithm considers levee crests with a convex profile (Fig. 3g–h). With flatter slopes in the coarser data, profile convexity is also closer to 0.

(v) Planform curvature

Finally, planform curvature ($plan_c$) is the second derivative of elevation orthogonal to the direction of maximum slope ($S'$), as defined in Eq. 6. This direction follows the levee crest, where very little change in slope is expected (Fig. 2f).

$$plan_c = \frac{dS'}{dxy} \tag{6}$$

From Fig. 3i–j, it is evident that levees in the NLD have no or very little planform curvature within a narrow tolerance.




344

345 **Figure 3.** Histograms showing the geomorphometric parameter thresholds sampled from

346 the USACE levee data at ⅓ (left-hand side) and ⅑ arc sec (right-hand side): **(a)(b)**

relative elevation; **(c)(d)** slope; **(e)(f)** aspect difference; **(g)(h)** profile curvature; and **(i)(j)** planform curvature.

Fig. 2g demonstrates an example of how the algorithm functions for a small section of the ⅑ arc sec DEM shown, with higher values of extraction rate (ε) along the known levee crests than for other floodplain features. The algorithm was run for the contiguous US at ⅓ arc sec resolution and at ⅑ arc sec resolution where available to re-generate a 1 arc sec NED-based DEM with these hydraulically-important features preserved during the coarsening process. The 1 arc sec "defended" DEM built with the feature preservation algorithm run with ⅑ arc sec data is validated against surveyed levee crest heights in the California Levee Database (CLD) provided by the State of California Department of Water Resources. The CLD contains geodetically surveyed crest elevations referenced to the NAVD 88 datum for roughly 7000 km of Californian levees. Concurrently, the "undefended" 1 arc sec DEM using standard bilinear resampling and no additional algorithmic consideration of levee-like features is benchmarked against the CLD to chart the improvement in DEM crest elevation representation when the new method is employed.

The continental-scale hydrodynamic model Fathom-US was re-run with a new 1 arc sec DEM built with algorithmic output of the ⅓ arc sec data, owing to the seamless coverage of this higher resolution variant. Performance when run with the original (undefended) DEM is compared to that when simulated on the new DEM, using local flood maps built by the Iowa Flood Center (IFC) as a benchmark. The IFC, established by the State of Iowa, is charged with producing and sharing inundation maps for the purposes of flood research, mitigation, prediction and insurance (Krajewski et al., 2017; Chen et al., 2017; Horna-Muñoz and Constantinescu, 2018). Their maps consist of two modelling strategies: (i) complete-coverage statewide maps

built with 1D HEC-RAS models to underpin National Flood Insurance Program (NFIP) rate setting and (ii) more detailed urban flood maps at select locations, mostly simulated using coupled 1D/2D models. Both IFC modelling approaches are driven with multiple AEP flows calculated from USGS gauging stations or using standard USGS regression equations for ungauged streams. For the statewide maps, water surface profiles generated via step backwater calculations using the AEP discharge and surveyed or lidar-derived channel cross-sections are intersected with 1 m resolution lidar-derived DEMs. The detailed urban flood models utilize surveyed channel bathymetry merged with lidar elevation data, representing both rivers and floodplains. Hydraulic structures such as levees, weirs and bridges are surveyed and stitched into either the 1D channel representation or the 2D DEM. The 1D/2D HEC-RAS or MIKE FLOOD models, run at 10–20 m resolution, are calibrated to flow observations, water surface profiles or high water marks where available. Further details on IFC flood mapping procedures can be found in Gilles et al. (2012). Comparing Fathom-US output to both IFC modelling strategies permits performance benchmarking to wide-area studies which employ cruder representations of flow physics, yet are built with accurate local data, and also to extremely high-quality, engineering-grade inundation models whose high data requirements result in prohibitively high financial expense to produce them at larger scales. Should the large-scale model with improved defense representation achieve similar realizations of flood extent across different AEPs, it will be a vindication of the methodology presented.

The testing consists of the employment of ubiquitous binary pattern fit statistics (e.g. Aronica et al., 2002; Sampson et al., 2015; Werner et al., 2005; Wing et al., 2017):

$$\mathrm{HR} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1} \tag{7}$$

$$\mathrm{FAR} = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \quad (8)$$

$$\mathrm{CSI} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \quad (9)$$

$$\mathrm{EB} = \frac{M_1 B_0}{M_0 B_1 + M_1 B_0} \quad (10)$$

where *M* and *B* describe pixels of the model being tested (Fathom-US) and those in the benchmark data (IFC) respectively and subscripts 1 and 0 indicate whether a pixel is wet or dry in each model respectively. Hit Rate (HR; Eq. 7; optimum = 1) is a measure of model underprediction, penalizing only "misses" ($M_0B_1$). The False Alarm Ratio (FAR; Eq. 8; optimum = 0) measures model overprediction, where "false alarms" are exclusively penalized ($M_1B_0$). Critical Success Index (CSI; Eq. 9; optimum = 1) is the most discriminatory measure, penalizing both under- and over-prediction. Finally, the Error Bias (EB; Eq. 10; optimum = 0.5) metric is slightly different to that employed in previous studies. It still indicates the overall balance of under- and over-prediction, but ensures the magnitude of the deviation from the optimum in either direction is consistent: EB > 0.5 is overprediction; EB = 0.5 is unbiased; EB < 0.5 is underprediction.

To further discriminate between the IFC and Fathom models, we employ exposure-weighted metrics in a similar vein to those proposed by Pappenberger et al. (2007). Eqs. 7–10 are applied in the same way, but $M_xB_x$ are population counts within such areas rather than pixel counts (e.g. $M_1B_1$ is the total number of people within areas the model and benchmark data agree is inundated). Population counts are derived from the demographic map of the US Environmental Protection Agency EnviroAtlas program, where census block counts are dasymetrically downscaled to 30 m pixels based on land-use and slope. While the pixel-count metrics indicate

wide-area physical modelling ability where correctly modelling relatively simple phenomena (e.g. capturing river channels and flooding in the large portions of undeveloped floodplain) is rewarded, the population-count metrics indicate model performance in the most important (and difficult to model) areas in a risk context.

To demonstrate transferability of this method to other geographic regions, indicate the effect of it on large-scale models where similar data are available, and perhaps illustrate a use case in smaller scale studies also, the extraction algorithm was run over ⅓ arc sec resolution lidar data (provided at 2 m resolution, but resampled to ⅓ arc sec) of the Po River floodplain in northern Italy. Such a test case enables greater isolation of the effect of this new levee representation method in a hydraulic model, since crucial flood model components – a seamless lidar DEM, surveyed channel bathymetry and hydrography, and well-constrained inflow boundary conditions – are substantially more accurate than in the larger scale, AEP-simulating US model. The ~350 km stretch of the middle–lower Po considered is a large alluvial floodplain bounded by embankments containing a system of minor levees within. The model is driven with boundary conditions from an historical flood event in October 2000 using the same computational hydraulic engine as in Fathom-US, but executed fully in 2D (no subgrid channels). Instead, surveyed channel bathymetry is burnt directly into the DEM, thus representing rivers as supra-grid features. The data associated with this model of the Po has been used in a number of hydraulic studies (e.g. Castellarin et al., 2009, 2011a, 2011b; Di Baldassarre et al., 2010; Domeneghetti et al., 2015; Schumann et al., 2010). For this test case, three different DEMs are employed whilst all other elements of the model structure are held constant:

437 (i) manual-defended (MD): the raw DEM is resampled to 1 arc sec resolution and
438 manually digitized levee crest centerlines are used to elevate corresponding pixels
439 to the ⅓ arc sec elevation value;

440 (ii) automated-defended (AD): to emulate large-scale model structures, where local
441 levee data is mostly unavailable, the automated levee extraction algorithm is run
442 over the raw ⅓ arc sec DEM, preserving the elevations of relevant pixels during
443 the resampling to 1 arc sec resolution;

444 (iii) undefended: no representation of flood defenses except those which are inherent
445 in the resampled 1 arc sec DEM; while some residual representation of flood
446 defenses may be present, albeit dampened, no additional enhancement beyond
447 their coarsened representation is performed.

448 The methodology presented here would be validated if levees hold (i.e. are not
449 overtopped by floodwaters) in the AD (ii) DEM where they also hold in the MD (i) DEM while
450 overtopping in the undefended (iii) DEM. While the manual handling of flood defenses considers
451 exclusively formal levee structures, the automatic method extracts all levee-like features
452 regardless of their anthropogenic classification. This means the elevation of more informal
453 features, such as small berms, river banks, roads, undocumented (often minor) dykes and other
454 natural features, is captured, while remaining "smoothed out" of the MD DEM. As such, we can
455 further examine the hydraulic effect (e.g. on flood peak attenuation) of accounting for these
456 features in the AD vs. MD DEM using observations of flow and high water during the event
457 considered.

458

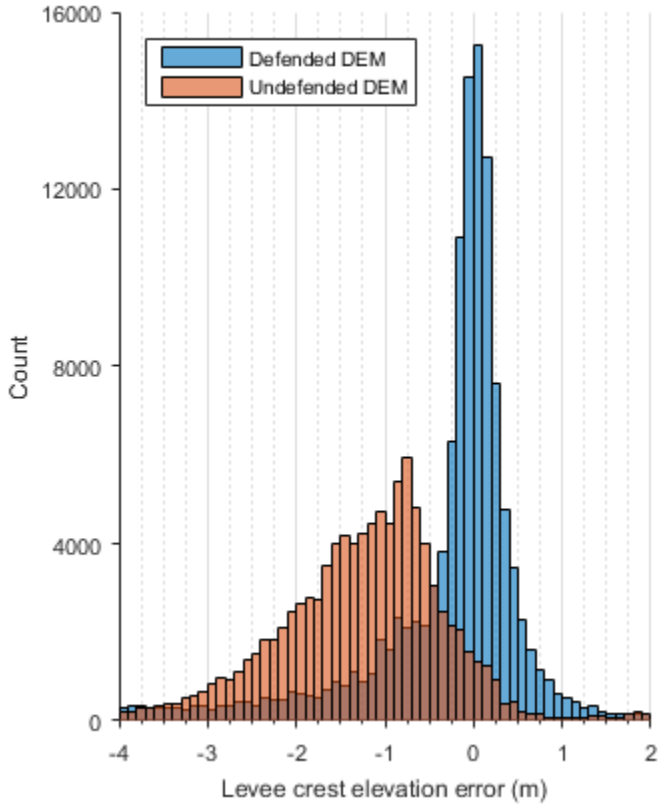
## 3 Results and Discussion

The algorithm was executed across the entire contiguous US at ⅓ arc sec (~10 m) resolution, at ⅑ arc sec (~3 m) resolution where available in the US, and over the Po floodplain at ⅑ arc sec (~3 m) resolution to ultimately produce new "defended" 1 arc sec (~30 m) DEMs for the US and the Po. At ⅓ arc sec resolution, a single parameter threshold set takes, on average, 0.99 s to run for a 1 x 1° (approx. 100 x 100 km) tile on a single 2.30 GHz Intel Xeon E5-2650 core. For the same size tile at ⅑ arc sec resolution, a single parameter threshold set takes 2.54 s to run on average. Given each parameter threshold set simulation is independent, the algorithm lends itself ideally to parallelization. Executing a hydrodynamic model at this scale and either resolution would require many orders of magnitude greater simulation time. Thus, preserving any relevant high-resolution information during the necessary coarsening of the DEM is computationally plausible and should enable improvements in model performance: particularly for lower-magnitude, higher-frequency flooding which is very sensitive to small-scale topographic features. This section will examine whether this is the case.

### 3.1 Levee crest elevation validation: California, US

The 1 arc sec DEM built with consideration of levee-like features at ⅑ arc sec and an ε threshold of 300 (see section 3.2.1 for rationale) is compared to geodetic surveys of levee crest heights in the California Levee Database. The results of this benchmarking are shown (in blue) in Fig. 3, where positive (negative) errors indicate DEM overprediction (underprediction) of the CLD crest height. Crest elevation errors appear broadly normally distributed around a central tendency close to zero. The defended DEM exhibiting a median error of –0.05 m means that, on average, crest elevations are underpredicted by 5 cm. Median absolute error (meaning positive and negative errors do not cancel out) comes to 0.26 m. Errors in the undefended 1 arc sec DEM built

482 only by resampling the ⅓ arc sec data are shown in orange. This DEM has a clear
483 underpredictive bias, with a median error of –1.16 m.




484

485 **Figure 4.** Histograms showing the distribution of elevation errors when benchmarking
486 the defended (blue) and undefended (orange) 1 arc sec DEMs against levee crest
487 elevations in the California Levee Database.

488

489 It is evident, then, that the smoothing effect of standard resampling approaches result in
490 substantial truncation of levee crest heights. The upshot of this is that flood inundation models
491 which have no *post hoc* techniques in the DEM processing for ameliorating this issue will
492 overtop levees and inundate areas they otherwise protect at shallower water depths than in

reality. Meanwhile, the new approach presented here appears to be highly effective in the preservation of crest heights; reducing average crest elevation errors to almost zero. That said, there is a hint of negative skewness in the defended DEM (mean error of –0.37 m), with some levee crests still being substantially underestimated. This illustrates the inherent challenges and inevitable smoothing effect of constructing DEMs (even at high resolution) where single elevation values often represent heterogenous grid cells. Some of the more extreme errors in both DEMs are likely where the NED is not built with lidar data.

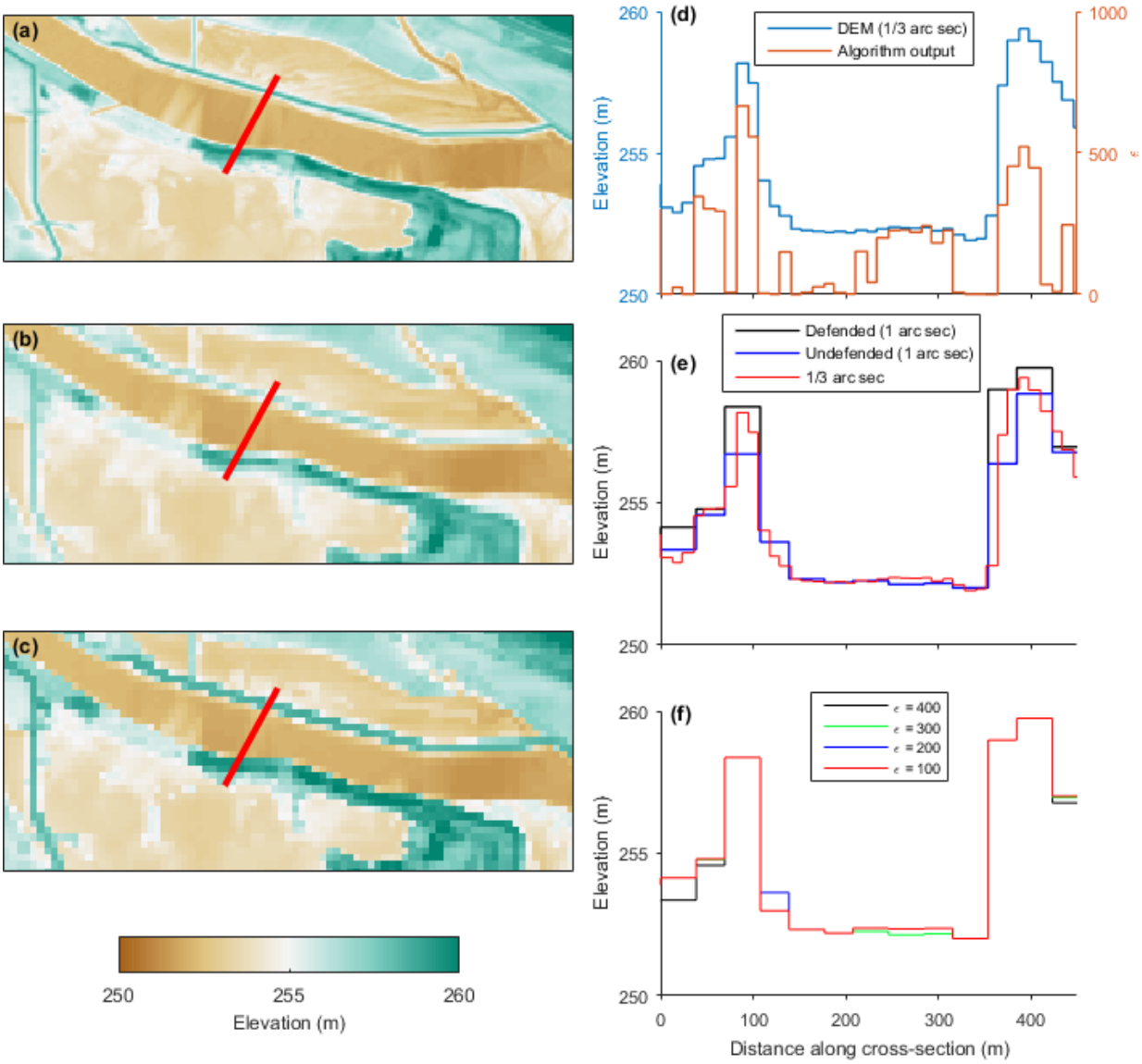
### 3.2 Inter-scale model comparison: Iowa, US

#### 3.2.1 Elevation model building and parameterization

The 1 arc sec DEMs employed in the Iowa analysis, both derived from the ⅓ arc sec NED in this instance, are exemplified in Fig. 5. It is evident that in this particular area of Iowa, a lack of consideration of levees results in a diminution of their elevation upon resampling to a coarser resolution: particularly on the northern bank of Cedar River (compare Fig. 5a and 5b). In the DEM generated via the levee detection algorithm, the levees are represented to their full effect (compare Fig. 5a and 5c). Fig. 5e provides a lateral view of this: both the undefended and defended DEMs are identical, except for their representation of the two levees either side of the river channel. The defended DEM has elevated the relevant pixels to more closely match the peaks captured in the ⅓ arc sec data. Fig. 5d plots the ⅓ arc sec elevation data again, but with the corresponding algorithm output. It is clear that levee crests are afforded higher values of ε than other features, but in some instances extraneous objects (e.g. artefacts in the DEM representation of the channel bed, points at deceleration of the levee slope) have notable ε values. Even if a

515 threshold of ε is employed which transpires to capture these objects however, the effect of this on
516 the DEM is minimal, demonstrating the tendency of this method to 'fail to safety'. Fig. 5f
517 illustrates this for different thresholds of ε: the resultant DEM scarcely changes in any
518 meaningful way when examining a range of suitable values. ε thresholds greater than 400 tend to
519 produce DEMs with broken levee centerlines and ignore more modest, yet potentially crucial,
520 levee-like features. Based on numerous visual inspections of the data across the US, an ε
521 threshold of 300 is identified to be suitable. Fig. 5f indicates that so long as the algorithm's
522 geomorphometric parameters are considered to some extent, DEM generation appears to have
523 little sensitivity to the choice of threshold. Evidencing this across a much larger scale, the
524 median non-zero difference between an $\varepsilon_{100}$ and $\varepsilon_{300}$ Iowa DEM is 0.12 m. In other words, pixels
525 where $100 \leq \varepsilon < 300$ are considered cause an average elevation increase of 12 cm compared to an
526 instance where they are not considered at all. This is well within the errors in the raw DEM itself
527 (Gesch et al., 2014).

528




529

530 **Figure 5.** An example of the algorithm's functionality of a 1 x 3 km area along Cedar
531 river in Waterloo, IA. **(a)** shows the baseline high-resolution (⅓ arc sec) DEM; **(b)** shows
532 an "undefended" DEM, resampled to 1 arc sec with no consideration of levees; **(c)** shows
533 a "defended" DEM generated by the algorithm presented here, where the ε threshold is
534 300. The red line in **(a)** – **(c)** is the cross-section for which data is shown in **(d)** – **(f)**
535 (increasing distance along the *x* axis relates to NE – SW movement along the cross-
536 section). **(d)** shows the elevation in the ⅓ arc sec DEM alongside raw algorithm output of

values of ε for each corresponding pixel; **(e)** shows the cross-sections of elevation values for each of the DEMs in **(a)** – **(c)**; **(f)** shows the cross-section of elevation values for when different ε threshold values are used.

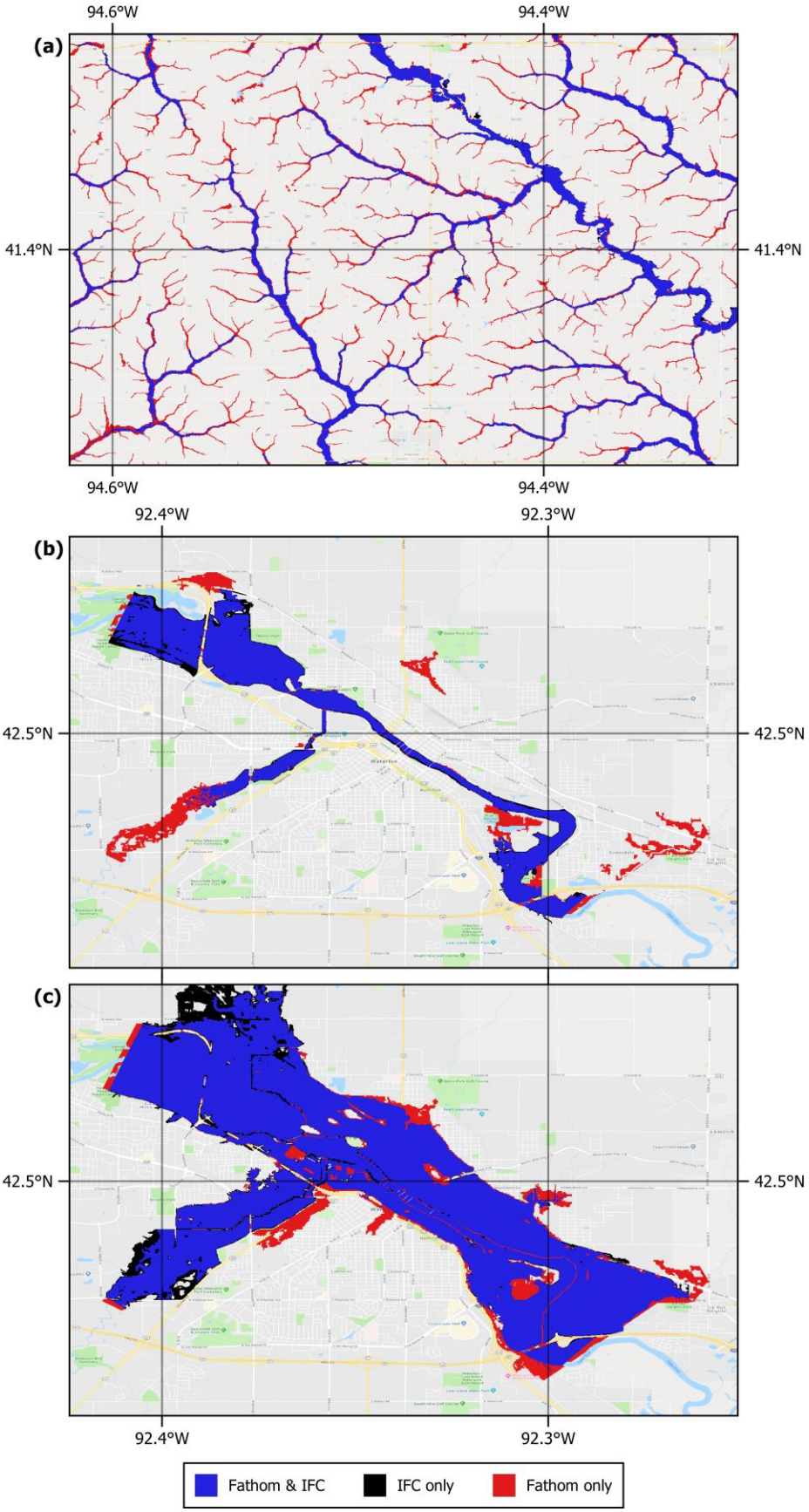
### 3.2.2 Statewide model comparison

Firstly, the results of the comparison to lower-quality, wide-area flood maps of the entire state are shown in Table 1. Generally, the two models show good correspondence with between 82% and 96% of the IFC floodplain captured by Fathom-US. Accounting for false alarms, which range between 15% and 27% of correctly identified pixels, the Fathom model attains a 63 – 82% fit to the IFC data (based on CSIs). Error biases indicate that 64 – 80% of incorrect pixels are false alarms. There is a clear trend in performance with AEP. As the floods get larger (AEP reduces; a lower-frequency, higher-magnitude flood is modelled), the fit to IFC data increases. This is perhaps unsurprising, as larger floods are often simply constrained by large-scale topographic features, while small floods are much more sensitive to the control exerted by small-scale topographic features. Differences due to boundary condition derivation on ungauged streams will also be more evident for smaller floods. Even slight differences in the definition of a 10% AEP streamflow can result in dramatically different flood extents since low-gradient floodplain land is still available for inundation. Differences in the 0.2% streamflow, so long as it is large enough to fill the valley, will make negligible difference to resultant flood extents. Stephens et al. (2014) also note that larger floods generally enjoy inflated CSI scores simply because there are more pixels to count as ‘hits’: misestimating the flood edge by a given distance is penalized more heavily for smaller floods. Performance disparities between large and small floods also arise through approximations of channel capacity. The IFC maps benefit from local

bathymetric surveys, while the Fathom-US model must approximate this since remote sensing of channel beds is not yet possible. Approximations are made based on drainage area and an assumed AEP bankfull discharge of 50%. In reality, this assumption will not hold, particularly for engineered waterways. Smaller floods will be much more sensitive to channel conveyance, as a greater proportion of the total discharge will be held in-channel compared to larger floods. Uncertainties relating to what the channel conveyance actually is will thus hold greater sway over the resultant extent of a smaller flood. One further phenomenon of note is the difficulty in isolating false alarms, noted also by Wing et al. (2017), due to differences in model domain. Through its nature as an automated, large-scale flood inundation model, all rivers of a certain drainage area (> 50 $km^2$) are modelled and the pluvial model simulates flooding in rivers smaller than this. The IFC data is an assemblage of local studies, which generally map rivers down to ~3 $km^2$ flow accumulation, though do not model every one: often for justified reasons, because such rivers are in uninhabited areas and there simply is no need to expend resources on modelling them. Large-scale models are agnostic of such priorities and therefore have total coverage. With no easy way of excluding these areas from the analysis, many pixels flagged as false alarms are truly in areas the IFC have not studied. See Fig. 6a for an example of this, where there is broad flood extent agreement in areas the IFC has modelled but non-genuine false alarms in unmodelled-by-IFC headwater areas. Thus, while the error bias scores indicate a relatively high tendency towards overprediction, metrics accounting for false alarms (FAR, CSI and EB) should be viewed in light of the limitations of the benchmark dataset.

| Annual Exceedance Probability | Hit Rate | False Alarm Ratio | Critical Success Index | Error Bias |
|---|---|---|---|---|
| 10% | 0.82 | 0.27 | 0.63 | 0.64 |
| 4% | 0.88 | 0.22 | 0.71 | 0.68 |
| 2% | 0.92 | 0.20 | 0.75 | 0.73 |
| 1% | 0.93 | 0.18 | 0.78 | 0.75 |
| 0.5% | 0.95 | 0.18 | 0.79 | 0.79 |
| 0.2% | 0.96 | 0.15 | 0.82 | 0.80 |

581 **Table 1.** Pixel-count test scores, comparing the statewide IFC maps and Fathom-US run
582 with the new defended DEM.


(a)
94.6°W
94.4°W
41.4°N
(b)
92.4°W
92.3°W
42.5°N
(c)
Fathom & IFC
IFC only
Fathom only

583

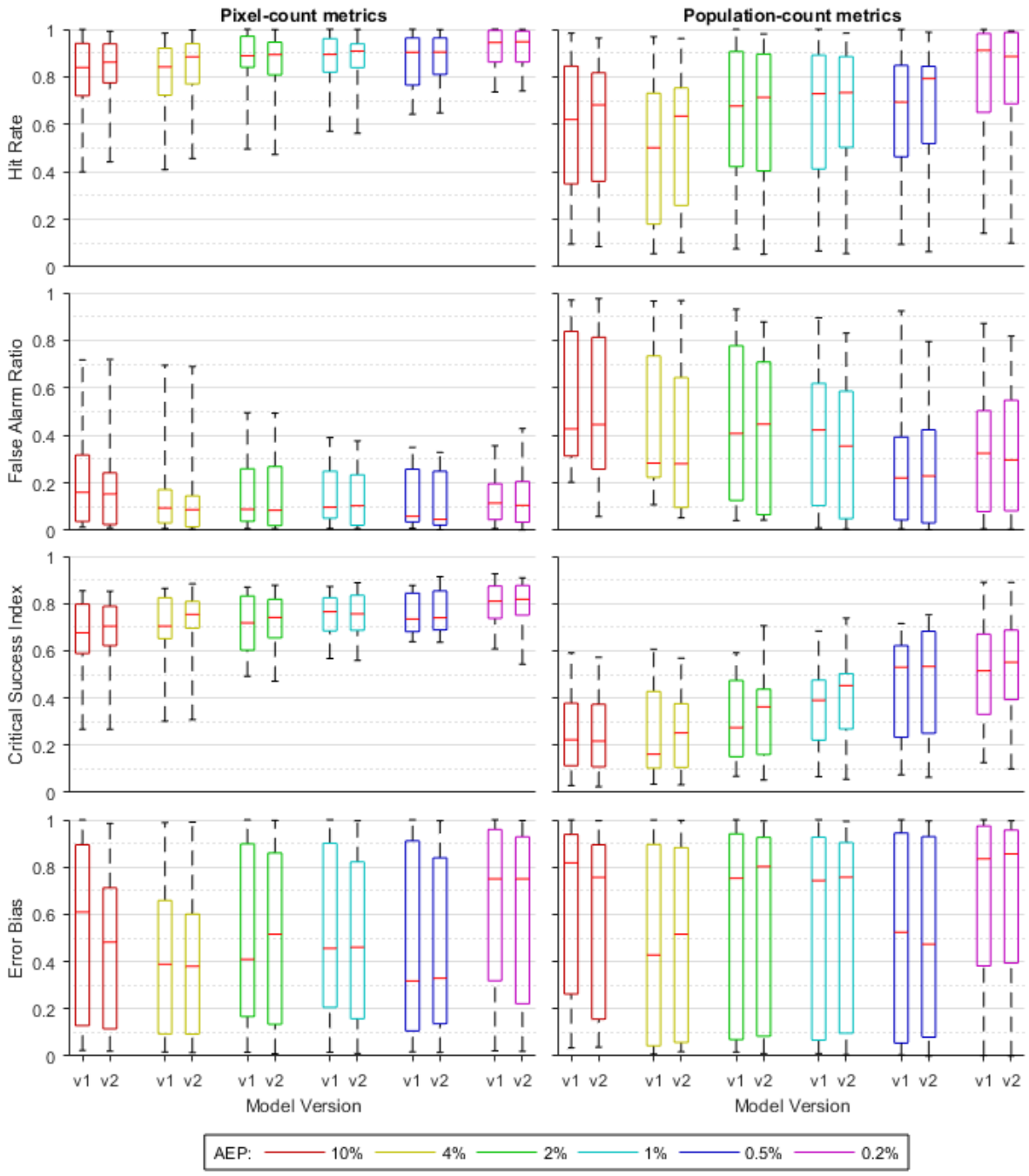
584 **Figure 6.** Example of the model intercomparison between Fathom-US and the IFC's **(a)**
585 statewide models of the 1% AEP flood for an area of Iowa in between Des Moines and
586 Omaha; urban model at Waterloo of the **(b)** 10% AEP flood and **(c)** 0.2% AEP flood.

587

588 3.2.3 Urban model comparison

589 With a more manageable number of high-quality IFC urban flood maps, model domains
590 and the rivers they include are easier to isolate manually; enabling a more faithful
591 intercomparison. However, there are still instances where the IFC have modelled a single river
592 flowing through the urban center, while Fathom-US also simulated flooding on its tributaries that
593 fall within the specified domain. For larger floods, areas around these tributaries are inundated
594 by the main river and require consideration. For smaller floods, these areas are unaffected by the
595 IFC-modelled channel, yet inundation arises from overtopping in the (unmodelled by IFC,
596 modelled by Fathom-US) tributary. See Fig. 6 for an example at the Waterloo study site. Some
597 areas flagged as false alarms for the 10% AEP flood (red in Fig. 6b) are not connected to the
598 IFC-modelled Cedar River, but around IFC-unmodelled Black Hawk Creek (SW), Virden Creek
599 (NE) and Elk Run (SE). The reason these inundated areas are considered (e.g. only a small
600 portion of Fathom-US flooding in Virden Creek) is because the 0.2% AEP flood (Fig. 6c) arising
601 from the Cedar River overtopping extends to these areas and so falls within the bounds of the
602 study area. In reality, the 10% AEP flood appears a close match along the Cedar River at
603 Waterloo but its CSI (and especially the population-weighted CSI) will not be a true reflection of
604 this. Thus, precise quantification of false alarms still remains elusive in some cases. The
605 comparison involved 27 individual studies encompassing the main urban areas within Iowa:
606 Ames, Cedar Falls, Cedar Rapids, Charles City, Clarksville, Columbus Junction, Des Moines,

Elkader, Fort Dodge, Greene, Hills, Humboldt, Independence, Iowa City, Kalona, Manchester, Maquoketa, Mason City, Monticello, Ottumwa, Plainfield, Red Oak, Rock Rapids, Rock Valley, Spencer, Waterloo and Waverly.




611

612 **Figure 7.** Boxplots of the binary pattern metrics when comparing version of Fathom-US

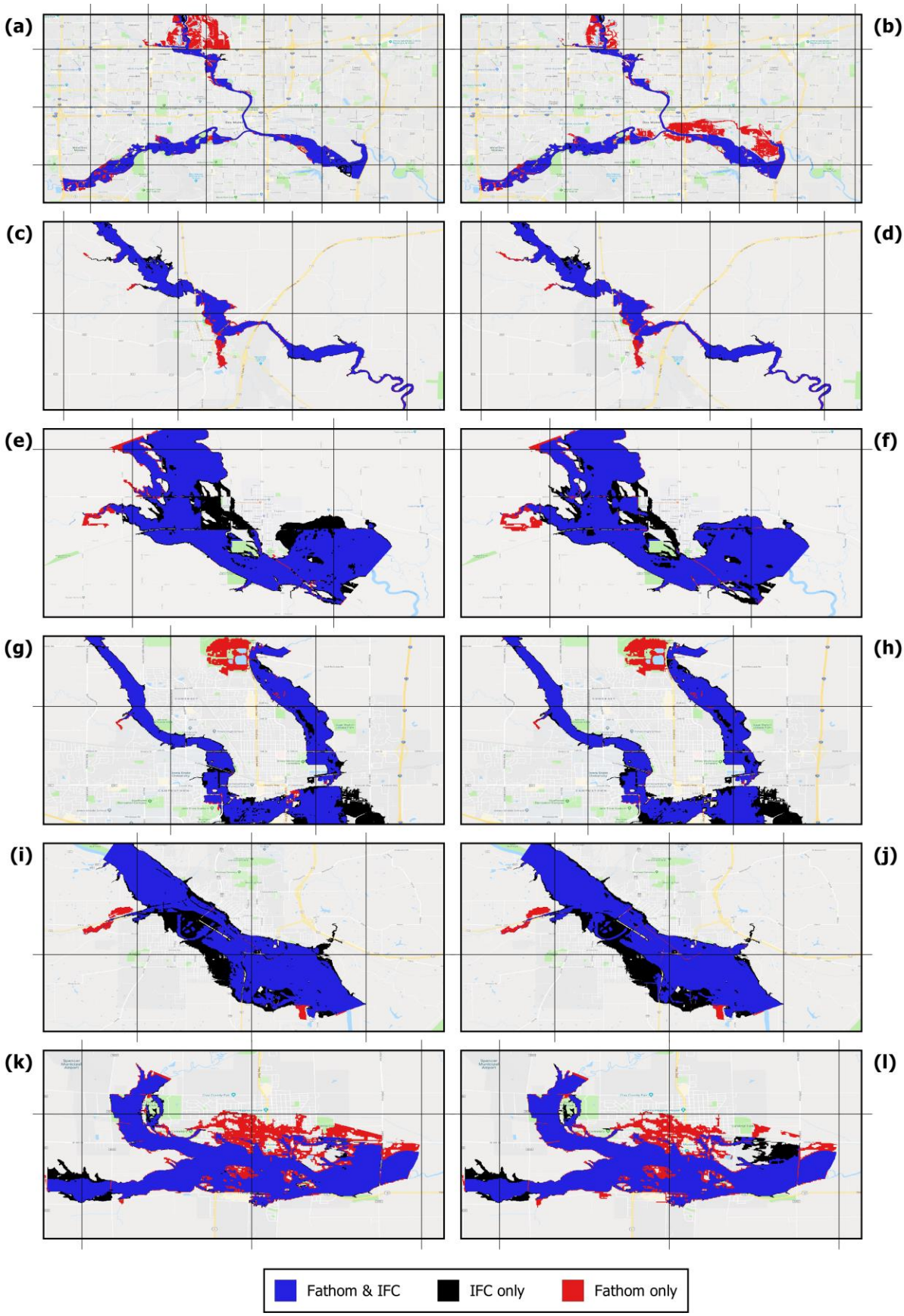
613 to the IFC models of 27 urban areas in Iowa. The left-hand side shows test scores using

raw pixel counts in Eqs. 7–10. The right-hand side shows test scores using population counts in Eqs. 7–10. The left boxplot (v1) within each of the six AEP groupings is the original version of Fathom-US presented in Wing et al. (2017), while the right boxplot (v2) in each group is the Fathom-US model when executed using the new defended DEM presented here.

Fig. 7 gives an overview of the intercomparison results, where both v2 (defended DEM presented in this paper) and v1 (standard DEM resampling) iterations of Fathom-US are benchmarked against the IFC models. Selected examples are shown as maps in Fig. 8, where performance has increased in Ottumwa (Fig. 8i–j), remained broadly constant at Monticello (Fig. 8c–d), and slightly decreased at Clarksville (Fig. 8e–f). Based on common pixel-count metrics, performance of the new v2 model is generally high. Median hit rates range from 86% at the 10% AEP flood to 95% at the 0.2% AEP flood (all AEPs experience a maximum HR of > 99%). Median CSI at the 10% AEP is 0.69, up to 0.80 for the 0.2% AEP. Maximum CSIs for each AEP flood range from 0.85 (10% AEP) to 0.91 (1% AEP). False alarms are generally at ~10% of correct pixels in magnitude and tend towards accounting for 40–60% of incorrect pixels across all AEPs. However, it is evident that increases in performance based on simulating over a standard DEM are either minimal or non-existent. This is because the bulk of the floodplain has essentially remained unchanged between the two models, clouding small-scale and important changes to inundation extent by the defended DEM. Viewing population-weighted metrics goes some way in addressing this issue by being much more sensitive to small changes in the flood edge in inhabited areas, but even these scores are generally dominated by inundation in areas where the two DEMs are virtually identical or where baseline DEM representation of important

features is already adequate. Even so, increases in median population-count CSIs from v1 to v2 are quite marked: 0.20 to 0.22 (10% AEP); 0.14 to 0.25 (4% AEP); 0.23 to 0.36 (2% AEP); 0.33 to 0.45 (1% AEP); 0.50 to 0.53 (0.5% AEP); 0.50 to 0.55 (0.2% AEP). For instance, the 1% AEP flood in Ames (Fig. 8g–h) has a muted pixel-based CSI improvement from v1 (0.70) to v2 (0.72) but a relatively large increase in population-weighted CSI (0.39 to 0.47). This suggests model performance is increasing in the very areas we are most interested in (from a risk-based perspective). Interestingly, this does not appear to be driven by a reduction in false alarms (population-count FARs remain fairly constant across model versions) as might be expected, but by increases in the rate of correct identification of inundated areas (population-count HRs increase). This suggests that a more accurate representation of flow conveyance across the floodplain, driven by heightened representation of hydraulic structures, correctly inundates areas previously left dry by the model at a greater rate than preventing incorrectly inundated areas from flooding. On the other hand, the failure of this new method to reduce false alarms suggests there are other causes of overprediction inherent to large-scale models. Approximations to channel capacity may induce out-of-bank flow to occur more frequently than in reality. With the proportions of floodwater in-channel and on-floodplain skewed for a given AEP event, correctly represented flood defenses may incorrectly overtop regardless. In Fig. 9, it is evident that, in certain circumstances, this method of DEM construction actually exacerbates errors in other facets of the model. In Fig. 9a, important hydraulic features have been elevated by the levee detection algorithm around this stretch of the Missouri River, enabling their accurate representation in the inundation model. Yet, as indicated by the red line, the HydroSHEDS-derived channel network deviates substantially from the evident tributary branching eastwards from the Missouri (channel networks are derived from interpolating across the cell centers of 3

arc sec HydroSHEDS flow accumulation data). The flood walls (Fig. 9b) on the W and SE banks of the Missouri have held back flood waters, while the hydrography of the tributary has broken through the flood walls in the N and E of the figure. As bankfull flow was exceeded the floodwaters backed up against the levee, meaning it could not even enter the true channel or inundate the floodplain engineered to store and convey the water, and proceeded to flood an erstwhile defended area. A smoother DEM with a discontinuous representation of these features is more forgiving of hydrographic errors: at least by permitting water to inundate areas it would do in reality. There is an implication, then, that representing defenses in this way is overzealous in some situations given the errors posed by poor channel representation. Put another way, the use of channel hydrography defined by HydroSHEDS is inadequate in this sophisticated model setup. The integration of new data sources, such as the USGS National Hydrography Dataset based on the ⅓ arc sec NED (NHDPlus HR) or MERIT Hydro (Yamazaki et al., 2019), or DEM-based methods similar to the levee preservation method presented here which instead preserve hydraulically-relevant flow paths and channel networks (e.g. Moretti & Orlandini, 2018; Sangireddy et al., 2016), may ameliorate this issue if implemented in the Fathom-US model framework. Future research will explore this.




676

677 **Figure 8.** Examples of model intercomparison between the IFC's urban models and
678 Fathom-US v1 (right-hand side) and v2 (left-hand side). Each grid box is 180 arc sec (~6
679 km). The table below gives further information:

| Fig. 8 panel | Location | Fathom-US version | AEP | CSI (pixel-based) | CSI (population-based) |
|---|---|---|---|---|---|
| **(a)** | Des Moines | v2 | 10% | 0.71 | 0.43 |
| **(b)** | Des Moines | v1 | 10% | 0.65 | 0.31 |
| **(c)** | Monticello | v2 | 4% | 0.86 | 0.44 |
| **(d)** | Monticello | v1 | 4% | 0.84 | 0.43 |
| **(e)** | Clarksville | v2 | 2% | 0.78 | 0.10 |
| **(f)** | Clarksville | v1 | 2% | 0.81 | 0.09 |
| **(g)** | Ames | v2 | 1% | 0.72 | 0.47 |
| **(h)** | Ames | v1 | 1% | 0.70 | 0.39 |
| **(i)** | Ottumwa | v2 | 0.5% | 0.81 | 0.52 |
| **(j)** | Ottumwa | v1 | 0.5% | 0.74 | 0.26 |
| **(k)** | Spencer | v2 | 0.2% | 0.74 | 0.45 |
| **(l)** | Spencer | v1 | 0.2% | 0.76 | 0.51 |

680

681

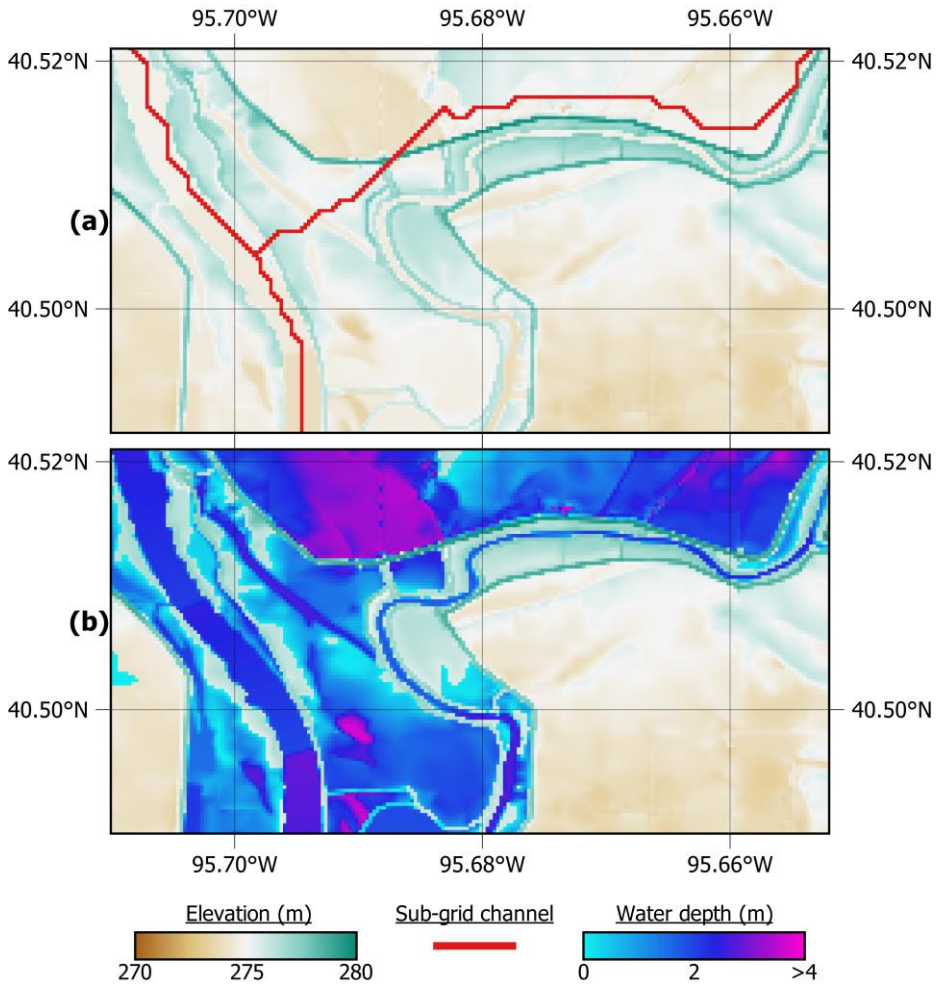



**Figure 9.** For a stretch of the Missouri River: **(a)** the defended DEM and sub-grid channel network derived from HydroSHEDS employed by Fathom-US, with **(b)** the resultant 2% AEP flood depth grid.

More broadly, the population-weighted metrics indicate there is a lot more work to be done to resolve the scale performance gap in inundation modelling than would be suggested by the pixel-based metrics. The pixel-based scores here and in Wing et al. (2017) suggest skill between local and large-scale models is converging, with CSIs approaching a ceiling given the fundamental hydraulic modelling constraints imposed by extreme flow characterization (Wing et

al., 2018). This makes charting model improvements to seemingly already within-error models difficult. However, while such models may appear to be within or approaching error on a pixel-count basis, the exposure-weighted metrics illustrate stark deviations in relevant inundation extent. An idea of what a high performance model should score in this context has little precedent though, so expectations (e.g. pixel-based CSI of > 0.75 amounting to excellent performance; Fleischmann et al., 2019) should be tempered as population-based metrics will not be inflated by a large number of 'easy hits'. Regardless, the updated Fathom-US model finding median correspondence with local IFC studies of 20 – 55% of exposed populace suggests there is still plenty of room for improvement.

It is also important to note that validating hydraulic models of this scale is notoriously difficult. Such models are not event-replicating, but instead simulate something unobservable in reality (static-AEP-in-space floods), and, even if they were, observations of real events are unavailable at a spatial scale to adequately interrogate them. This necessitates model–model intercomparisons as a pseudo-validation procedure. The engineering-grade IFC models, though treated as a benchmark, are themselves uncertain. They generally benefit from calibration, where model parameters are adjusted until the model resembles specified benchmark data (e.g. high water marks). The higher model skill attributed to local-scale compared to large-scale models, then, may be a product of this non-physical tuning rather than possessing a more fundamentally skillful model structure. Furthermore, in spite of accurate elevation data, channel hydrography and bathymetry, computational hydraulics, and defense information, they are impaired by the fundamental inability of all models to characterize extreme flows. Even for models fortunate enough to have boundary conditions informed by a river gauge, their short records and subjection to non-stationarity (particularly of land-use) result in multiple defensible

715 interpretations of given AEP flows (particularly less frequent ones). The justification, therefore,
716 for one model being 'validater' and the other being 'validatee' is weakened. This perhaps
717 explains traces of equifinality in the pixel-based comparisons, with improvements to model
718 performance between defended and undefended models virtually undetectable amidst insoluble
719 boundary condition uncertainty. See Smith et al. (2015), Villarini et al. (2009), and Blöschl et al.
720 (2013) for further discussion of discharge estimation errors in time and space. That said, the
721 lower quartile of pixel-based CSI scores for high frequency design floods noticeably increasing
722 from v1 to v2 does suggest poorer models are being brought closer to the performance ceiling.
723 Furthermore, conventional pixel-based metrics would consider Fathom-US a high-performance
724 model and it is clear that, on the whole, the representation of hydraulically important features by
725 the algorithm presented here has improved the model further. For sporadic, qualitative
726 evaluations, the updated model appears more behavioral than previously. In Fig. 8a, levee
727 representation in Des Moines has resulted in constrained 10% AEP flooding through the urban
728 center, which v1 struggled to do (Fig. 8b). Similarly in Waterloo, the 10% AEP flood was
729 constrained (Fig. 5b). Yet, it is evident that floodplain feature preservation is no panacea:
730 uncertainties relating to other model components inhibit substantial gains in performance. In Fig.
731 8k, overprediction is still rife at Spencer regardless of DEM representation owing to
732 disagreement over boundary conditions. The IFC 0.2% AEP model here used a peak discharge of
733 1265 $m^3s^{-1}$, while the input to Fathom-US was 1580 $m^3s^{-1}$. It's unsurprising, therefore, that
734 simply differing the method of DEM construction did not reduce the false positives evident in v1
735 (Fig. 8l).
736

737 ### 3.3 Observation-based validation: Po River, Italy

738 #### 3.3.1 Qualitative examination of levee overtopping

739 The 2000 flood event in the Po floodplain was large enough in magnitude to submerge
740 areas of the floodplain protected by the minor dyke system, but not so large that the major
741 embankments were overtopped (estimated AEP of 2%). The most obvious test of the three model
742 runs (manual-defended, automated-defended and undefended), therefore, is a simple evaluation
743 of any overtopping of the outer levees. The MD model, as expected, passes this test in
744 overtopping only minor internal levees (Fig. 10a–di). The AD model (ε threshold of 300 in light
745 of its functionality in section 3.2 and crest elevation accuracy in section 3.1) does this too: flow
746 is contained entirely within the floodplain bounded by the outer levees (Fig. 10a–dii). Crucially,
747 the undefended model run simulates an overtopping of these outer levees across the domain (Fig
748 10a–diii). Though a somewhat aggregated and narrow view of model performance, levees
749 holding in both defended model runs while overtopping in the undefended run demonstrates the
750 algorithm presented here represents flood defenses in a similar way to a laborious, manual,
751 unscalable method. The undefended model here is essentially a proxy for how existing large-
752 scale models operate: with no manual treatment of flood defenses possible, levee representation
753 remains diminished in the DEM and so floods that would be constrained in reality are not in the
754 model. It likely stands to reason that all typical large-scale models fail to adequately model the
755 Po River or any other defended river reaches globally in light of this analysis. Moreover, even
756 when built with lidar topography and continuous surveyed bathymetry, the undefended model
757 fails to simulate this event on the Po accurately since major outer embankments overtop. Since
758 they (correctly) did not overtop in the AD and MD models, this inaccuracy in inundation
759 simulation is attributable to the DEM resampling process. Large-scale models, which would

typically not benefit from even these data, are likely of little utility in modelling the Po. The specific topography of the basin – a levee-protected floodplain which is flooded only when the stable channel experiences significant stage – poses a difficult modelling challenge (Castellarin et al., 2011a). Even in the presence of complete levee inventories, implementation of these data within the model build is sensitive and non-trivial. Meanwhile, the automated method presented reproduces experiential expectations of levee behavior akin to local modelling strategies.

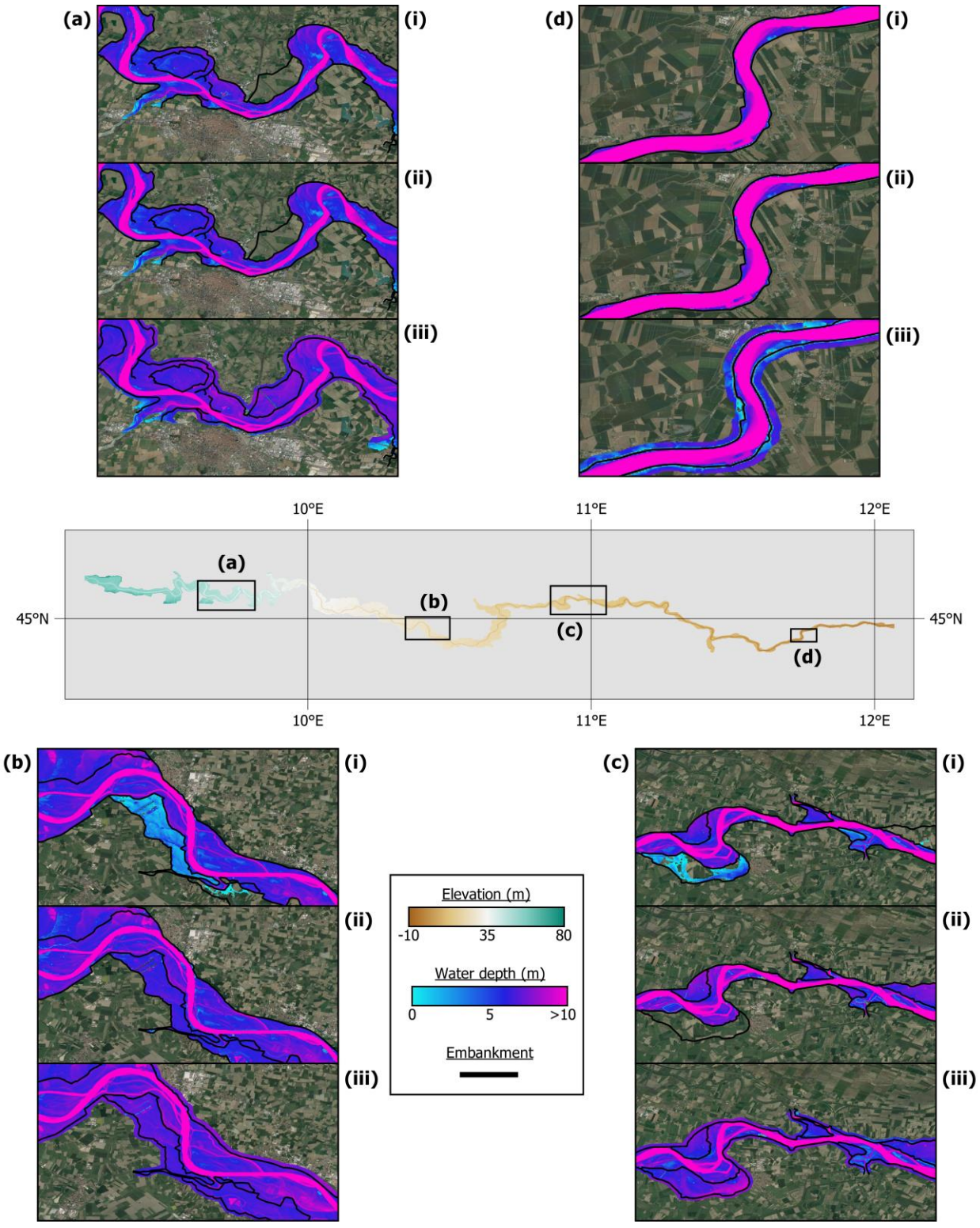



**Figure 10.** Four areas – **(a)** Piacenza, **(b)** Casalmaggiore, **(c)** just upstream of Ostiglia, **(d)** just downstream of Pontelagoscuro – of the simulated 2000 flood event along the Po

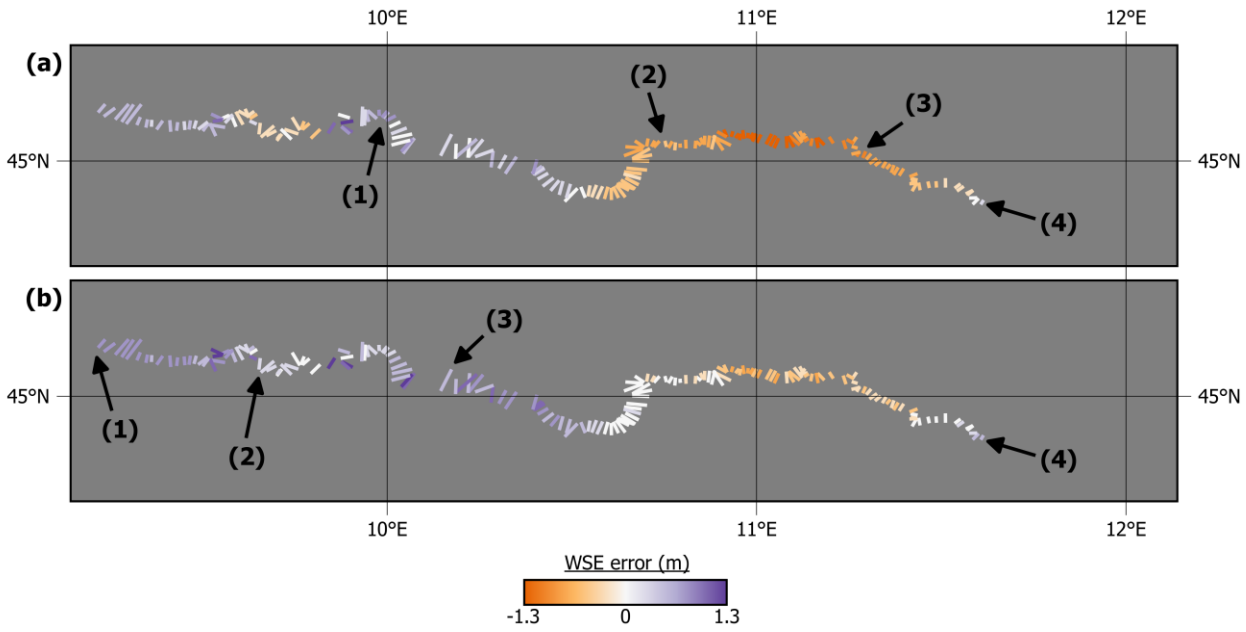
River for the three different DEMs: **(i)** manual-defended; **(ii)** automated-defended; **(iii)** undefended. The DEM only extends ~200 m from the outer embankments, making overtop visualization difficult. In all four example panels (and, indeed, across the entire domain) there is no outer levee overtopping in **(i)** and **(ii)**, while overtopping in **(iii)**.

### 3.3.1 Validation against high water marks

With the two simulations involving some treatment of defense structures performing similarly in terms of levee overtopping, we can further discriminate between the two models with some observational data. These take the form of post-event field surveys of maximum water surface elevation (WSE) for the 2000 flood event. In this analysis, we use 171 cross-sections of maximum WSE distributed across the length of the Po domain studied (Fig. 11). For the MD simulation (Fig. 11a), the Root Mean Squared Error (RMSE) is 0.61 m with a slight bias towards underprediction (mean error = – 0.25 m). With a DEM generated using the algorithm presented here (Fig. 11b), the RMSE comes to 0.46 m with a minor tendency towards overprediction (mean error = 0.15 m). These errors are commensurate with those reported when validating high-resolution, calibrated, local models built with quality topography, bathymetry and flow data (e.g. Mignot et al., 2006; Neal et al., 2009). Furthermore, errors in observed high water marks during post-event surveys are typically 0.3–0.5 m (Fewtrell et al., 2011; Horritt et al., 2010), meaning both models are within or approaching observational error. Both models employ distributed roughness values which are calibrated to the MD model, resulting in physically realistic drag coefficients similar to previous studies (e.g. Castellarin et al., 2011a): Manning's *n* in channels is set at 0.0320–0.0395 and floodplains at 0.0430–0.0950. Importantly, errors are lower in the

792 automated method of representing defenses than in the ubiquitous manual approach, in spite of
793 friction being calibrated to the latter.

794




795

796 **Figure 11.** Maps showing the spatial distribution of maximum water surface elevation
797 errors at each cross-section when comparing observations to **(a)** manual-defended and **(b)**
798 automated-defended model output. Positive errors (greys) indicate model overprediction;
799 negative errors (reds) indicate model underprediction. The direction of flow is west to
800 east. The numbers in **(a)** correspond to river gauges in Fig. 14. The numbers in **(b)** refer
801 to locations in Fig. 12 or are referred to in the text.

802

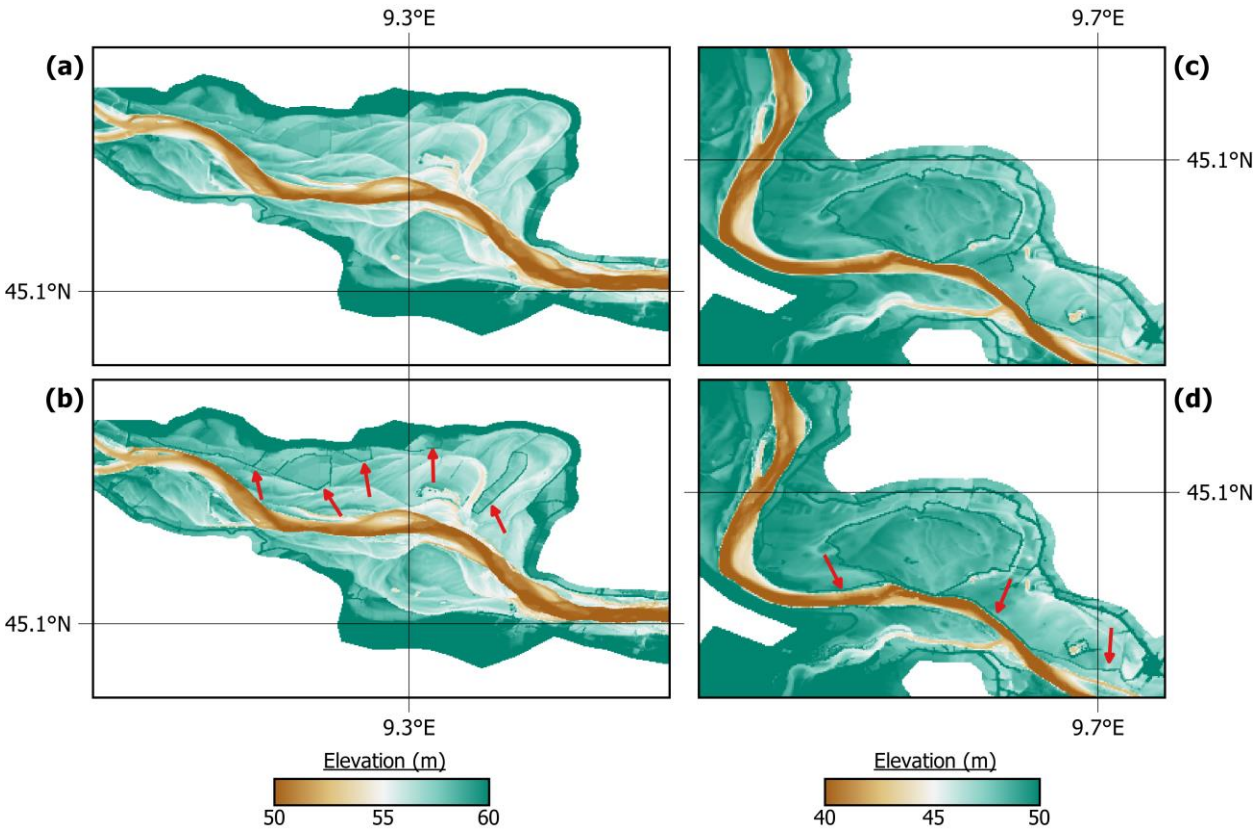



**Figure 12.** Maps highlighting the minor, yet crucial, MD and AD DEM differences (indicated by the red arrows). **(a)** and **(b)** are MD and AD DEMs respectively of the upstream domain boundary, at the confluence of the Po and Ticino rivers downstream of Pavia (location 1 in Fig. 11b). **(c)** and **(d)** are MD and AD DEMs respectively of an area just upstream of Piacenza (location 2 in Fig. 11b).

In both models, errors are generally low at the upstream (westerly) end of the domain, but performance begins to diverge at roughly 10.5°E, between Casalmaggiore (location 3 in Fig. 11b) and Pontelagoscuro (location 4 in Fig. 11b). The MD simulation becomes very underpredictive of maximum WSE here (up to 1.3 m at around 11°E), while performance remains high in the AD model. This suggests that the propagation of the flood wave is more

accurately represented in the latter simulation than the former. The likely reason for this is the method presented here not only preserves levee crest elevations in an efficient manner, but also represents other hydraulically important features to the extent that flood peaks across the domain are better captured than in models with a manual treatment of solely formal levee structures. Fig. 12 exhibits this for areas in the upstream model domain. Levee compartments near the domain entrance (Fig. 12a–b) were seemingly absent from the levee inventory used by the MD DEM, but were automatically detected for the AD model. Near Piacenza (Fig. 11c–d), levee-like features on the river bank are considered in the AD DEM, particularly on the northern banks, while remaining unresolved in the MD version. Maintaining river bank heights and elevating other micro-topographic features that remain sub-grid to the MD model resulted in faster conveyance of floodwaters downstream in the AD model, culminating in a more accurate representation of maximum WSE in the easterly (downstream) portion of the domain. Fig. 13 illustrates this assertion. The widespread areas of purple shading in Fig. 13a indicates such areas were inundated much earlier (6–12+ hours) in the MD model than the AD model. The flood wave in the AD model thus propagated further downstream, rather than being stored and conveyed more slowly out-of-bank in these areas. This is corroborated by the blue compartments of Fig. 13b, which show much deeper waters in the MD than AD simulation after 45 hours of model-time: a volumetric difference of ~2,450,000 $m^3$, despite identical boundary conditions. Out-of-bank flow or extensive floodplain water storage upstream that did not occur as early in the 2000 event (or the AD simulation) seemingly resulted in underprediction of maximum WSE downstream in the MD model. Representation of flood dynamics is less important in the upstream domain, since performance here is constrained by proximity to the in-flow points. This is perhaps why both models perform similarly west of 10.5°E. Further downstream, however, a realistic conveyance

of floodwaters begins to hold greater sway over performance (when comparing to WSE maxima).

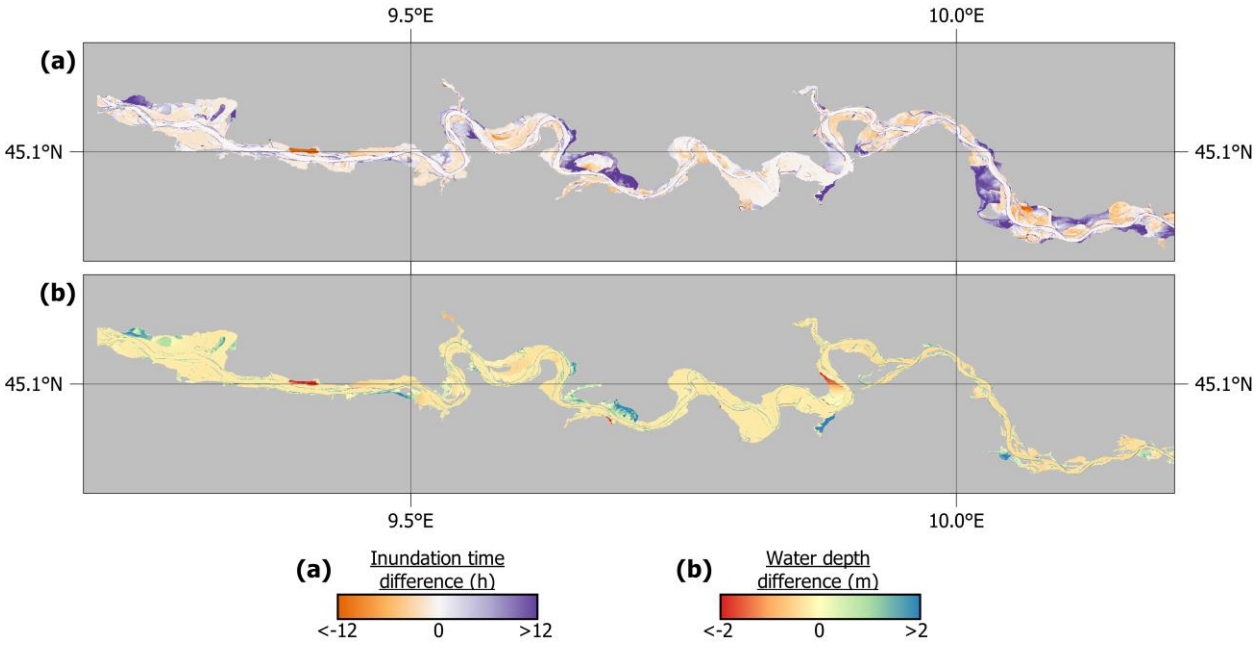



**Figure 13.** Maps of the upstream portion of the Po River floodplain showing: **(a)** time of inundation difference between the models, where a positive (negative) difference indicates earlier (later) inundation in the MD than AD model; and **(b)** water depth difference 45 hours into the simulations, where a positive (negative) difference indicates deeper (shallower) water in the MD than AD model.

### 3.3.2 Validation against river gauge data

This phenomenon is yet further evidenced by comparison to river gauge observations of water surface elevation. When comparing modelled WSE to those measured by the gauge at Cremona (upstream section of the domain; location 1 in Fig. 11a), both the MD and AD models perform similarly (Fig. 14a). The MD model is slightly underpredictive at Cremona, consistent

852 with the reasoning set out previously, while the AD model is marginally overpredictive here. It
853 should be noted that gauged observations of water level are highly accurate (< 1 – 2 cm),
854 providing suitably discriminatory validation data (Hamilton & Moore, 2012; McMillan et al.,
855 2012; Schmidt, 2002; van der Made, 1982). At Borgoforte (location 2 in Fig. 11a), the MD
856 model underpredicts the peak to a much greater magnitude (Fig. 14b). The AD model replicates
857 observations well here. Disparities on the rising limb are predominantly a function of model
858 spin-up, e.g., no water reaches Borgoforte until ~50 hours of model time; the Po would not be
859 empty before this time in reality. At Sermide (location 3 in Fig. 11a), the MD model is severely
860 underpredictive: misestimating peak WSE by ~1 m (Fig. 14c). The AD model underpredicts the
861 peak also, but only by ~0.5 m. The geometry reproduced by the AD approach seems to ensure a
862 more faithful reproduction of the channel–floodplain interaction, providing an attenuation more
863 similar to observations than with that obtained by the MD model. However, at the furthest
864 downstream gauge (Pontelagoscuro; location 4 in Fig. 11a), both models become slightly
865 overpredictive (Fig. 14d). Indeed, the MD model replicates observed WSE well at this gauge.
866 This is inconsistent with the previously supported interpretation of results whereby the MD
867 model fails to propagate the flood wave as faithfully as the AD model. The likely culprit here is
868 the horizontal resolution of the DEM. Pontelagoscuro is a location where the channel narrows to
869 150–200 m, with the large outer embankments only ~50 m from the river bank on either side.
870 The Po channel in this domain can be up to 500 m wide, bounded by up to 3 km of floodplain.
871 This is effectively a ‘bottleneck’, and so water surface elevation will be very sensitive to the total
872 storage volume afforded by the DEM. In such a narrow corridor, the horizontal resolution of the
873 DEM becomes more important in defining this. Fig. 14e is a cross-section of the channel and
874 floodplain at Pontelagoscuro. It is evident that both the MD and AD DEMs reduce the capacity

of the channel and floodplain here when compared to the high-resolution (⅓ arc sec) source data, simply as a function of grid resolution. Preserving peaks from the ⅓ arc sec data in the 1 arc sec DEM results in overrepresentation of these features in the horizontal plane. It is evident that there is less room for water storage in the coarser DEMs than would be present if the baseline ⅓ arc sec DEM was used. This is an important limitation to representing defenses in this way, but crucially is endemic to both the common-practice, engineering-grade, manual techniques and the automated approach presented here. The accurate replication of observed WSE by the MD model is likely incidental therefore: the result of the previously asserted underprediction of discharge combined with lower storage capacity. An overpredictive bias in the AD model at Pontelagoscuro is the probable product of more accurate modelled flow (higher than in the MD model) but with reduced storage volume available. Overall, therefore, there is consistency in the idea that the automated model not only performs similarly to manual methods by appropriately elevating flood defenses, but can outperform them in ensuring representation of other hydraulically important sub-grid scale features. At least, this is apparent on the Po, where representation of the minor embankment system is critical in the accurate replication of flood dynamics in the downstream portion of the domain. It should be noted that an MD approach can still obtain AD-like performance if these minor embankments are considered manually. In this test case, MD performance is an incidental product of the levee inventory containing only major levees. Should a modeler obtain and process extra information on these apparently critical floodplain features, the MD model would be indistinguishable from the AD model. The AD model, though, has the advantage of requiring no prior consideration of river dynamics in assessing which features to take note of. This suggests such an approach may offer benefits to local modelers, too, in improving model-build efficiency.

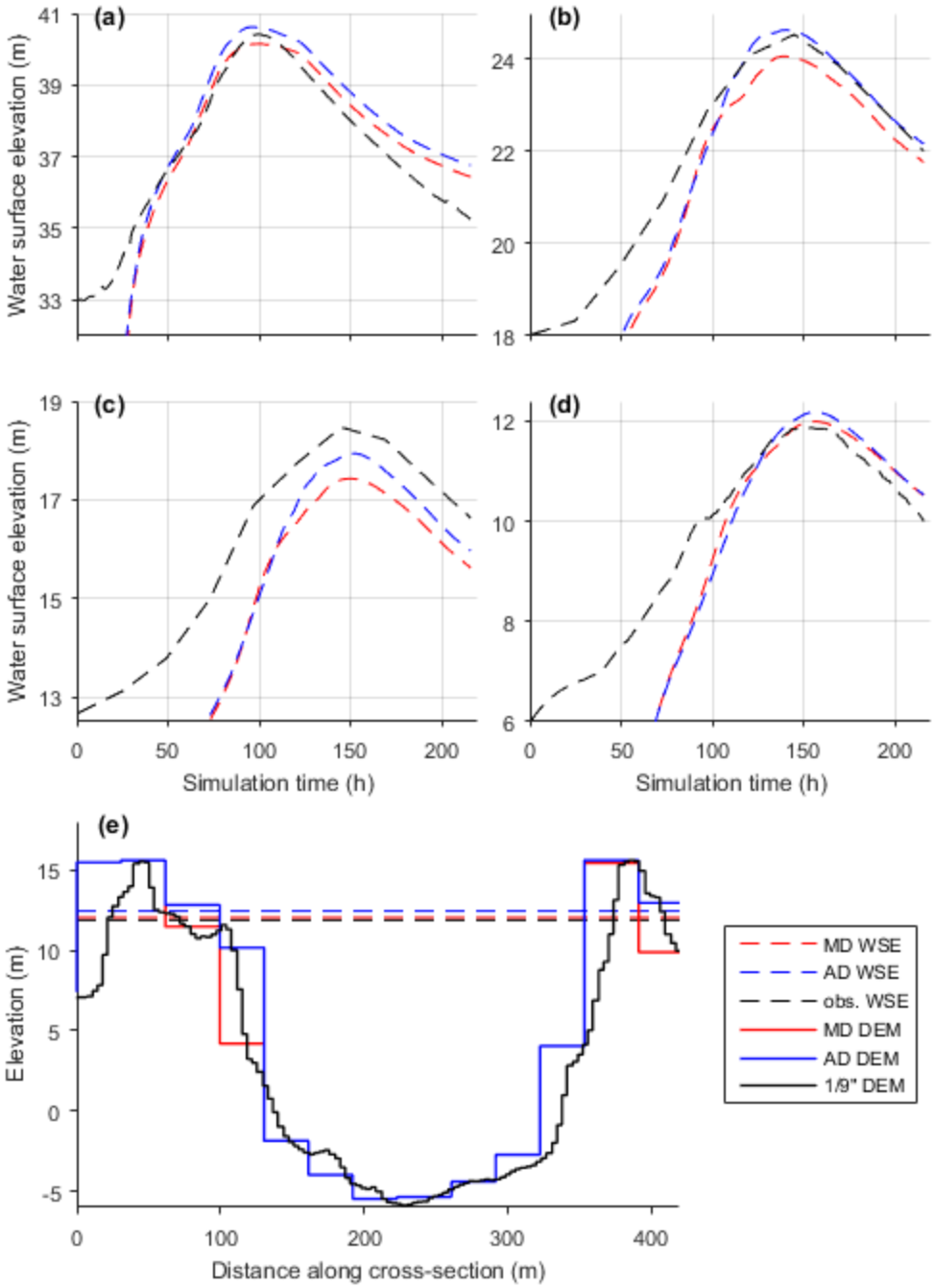



**Figure 14.** Comparison of MD and AD modelled water surface elevations to four river gauge observations of the 2000 event at: **(a)** Cremona (hourly data); **(b)** Borgoforte (daily data); **(c)** Sermide (daily data); **(d)** Pontelagoscuro (hourly data). Locations **(1)** – **(4)** in Fig. 11a correspond to gauges **(a)** – **(d)**. **(e)** is a cross-section of MD, AD and baseline ⅑

903 arc sec DEMs as well as MD, AD and observed water surface elevations at
904 Pontelagoscuro.

905

906 **4 Conclusions**

907 Overall, then, the new method of defense representation in DEMs presented here has
908 significant skill in accurately defining levee crest elevations. When validated against the
909 California Levee Database (section 3.1), crest heights were underestimated by the defended
910 DEM by 0.05 m on average, compared to 1.16 m by the undefended DEM. When incorporated
911 into a hydraulic modelling framework, the presented method of DEM building generally
912 improved flood extent and water level fits to model and observational benchmarks. Conventional
913 performance metrics remained high for the Iowa tests (section 3.2), with a trend towards
914 convergence with local-scale models. However, increases in skill versus an undefended model
915 were relatively muted, partly due to deficiencies in the comparison procedure itself, and inter-
916 scale model deviations in a risk context were substantial. The Po test case (section 3.3), though,
917 offers crucial evidence in support of the validity of this method when applied in conjunction with
918 more accurate data: (i) being a real, gauged flood event, boundary condition uncertainty is much
919 reduced relative to the theoretical design floods in the Iowa tests; (ii) complete channel
920 bathymetry and accurate hydrography offer considerable improvement in river representation
921 relative to the HydroSHEDS sub-grid channels employed in the US; (iii) total high-resolution
922 lidar coverage of the Po results in a more accurate DEM than that of Iowa when using the USGS
923 NED; and (iv) validatory benchmarks are better defined for the Po (observed high water marks,
924 gauged stage, no outer levee overtopping) than for Iowa (local-scale models).

This leads to a number of limitations in the application of this method to large-scale models. Firstly, it requires a high-quality, high-resolution baseline DEM that a modeler would typically coarsen. This effectively rules out the majority of the globe, where the best source of free elevation data remains corrected SRTM-based products, which have low vertical accuracy and are employed at native resolution. Applicable areas – those with wide-area lidar (or similar accuracy) coverage – are confined to North America, western Europe and parts of Australia. The method itself is grid- and neighborhood-scale dependent, requiring due consideration for different resolution source and simulation DEMs. Furthermore, the extraction rate threshold is difficult to define *a priori*. Based on visual inspections, we identify 300 to be appropriate but there is no formal justification for this. Derivation of a threshold from known defense locations is problematic, given this would unintentionally penalize the detection of hydraulically-important, non-levee features as false alarms, even if the reference defense inventory was complete. Despite this, we have shown the resultant DEM to be relatively insensitive to the threshold chosen. So long as a suitably low threshold is chosen to capture the features of interest, the effect of heightening extraneous features is negligible. Future research could seek to calibrate the extraction rate ($\varepsilon$) threshold to geodetic surveys of all hydraulically-relevant features or even treat $\varepsilon$ as an uncertain parameter in hydraulic model calibration studies: maximizing some measure of fit to a benchmark by varying the $\varepsilon$ threshold. Further, the method is agnostic towards the features it identifies and elevates; there is no judgement on their characteristics except height. The fragility of these features is thus unknown: they are not all engineered structures designed to hold back floodwaters, but could fail under such conditions. In which case, undefended model runs may still be useful to explore impacts should the elevated features not remain so during a flood. One final limitation, evident in the Po versus Iowa tests, is that other features of the

948 hydraulic model need to be of commensurate quality for any benefit to be felt. The method will
949 have no meaningful benefit for models utilizing hydraulic codes with poor physical
950 representation of flow, an over-coarsened DEM (e.g. > $10^2$ m resolution), inaccurately defined
951 boundary conditions and deficient representation of river channels.

952 Fathom-US generally performs well in Iowa, yet there is clear room for further
953 improvement. The Po model provides an indication of simulation quality should other model
954 components be improved. While improvements to extreme flow quantification are likely
955 impossible without denser river gauge networks and centuries of observations, large-scale
956 modelers can anticipate developments elsewhere. Increased coverage of high-quality elevation
957 data is eminently possible with relatively modest investments from global governments
958 (Schumann & Bates, 2018); updates to the USGS National Hydrography Dataset may supplant
959 the cruder global HydroSHEDS representation of river channels for the US and MERIT Hydro
960 shows promise at the global scale (Yamazaki et al., 2019); and emerging datasets detailing
961 fluvial geomorphological relationships (e.g. Frasson et al., 2019) from satellite data can be used
962 to constrain approximations to channel capacity, while NASA's imminent Surface Water and
963 Ocean Topography satellite mission (Biancamaria et al., 2016) may provide new opportunities
964 for estimating river bathymetry from space (e.g. Yoon et al., 2012; Pavelsky et al., 2014).
965 Meanwhile, this method may find applicability in smaller-scale studies given the improvements
966 to common local modelling strategies of the Po. Aside from being a more faithful DEM
967 representation of the floodplain, raw algorithm output can be used to identify areas requiring
968 further survey and automation provides important advantages over the laborious and non-trivial
969 process of manual burning.

970 Thus, the method is shown to provide skill in the detection of flood defenses and other
971 important hydraulic features in the absence of local information. It provides a crucial first step in
972 ensuring physically realistic flood defenses in large-scale models, moving away from the largely
973 unsubstantiated assumptions underpinning previous approaches. When it is a lack of defense
974 information that is the dominant source of error in a hydraulic model, the method presented here
975 offers considerable improvement.

976

977 **Acknowledgments**

978 The US Army Corps of Engineers National Levee Database is available from
979 https://levees.sec.usace.army.mil/. The US Geological Survey National Elevation Dataset can be
980 accessed from https://ned.usgs.gov/. The socio-economic data in Fig. 1 is available from the US
981 Census Bureau (https://www.census.gov/data.html) and the Multi-Resolution Land
982 Characteristics Consortium (https://www.mrlc.gov/). The authors thank Agenzia Interregionale
983 per il Fiume Po and Autorità di Bacino del Fiume Po for granting access to their data of the Po
984 River basin. They are also grateful to the State of California Department of Water Resources for
985 granting us access to the California Levee Database. The Iowa Flood Center benchmark
986 inundation maps are available from the Iowa Flood Information System
987 (https://ifis.iowafloodcenter.org/). The dasymetric population map of the US can be obtained
988 from the EPA Environmental Dataset Gateway (https://edg.epa.gov/). Fathom-US is a
989 proprietary model, but can be made available for academic research use (i.e. not for commercial,
990 policy or regulatory purposes) by contacting info@fathom.global. OEJW, PDB, JCN, CSS, AMS
991 and NQ are affiliated with Fathom, a flood analytics firm based in the UK. PDB was supported
992 by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research

Merit Award. JCN was supported by National Environmental Research Council grant NE/S006079/1. DWG, RG and WFK were supported by the Iowa Flood Center.

## References

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., and Feyen, L. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, 5, 171–182. doi:10.1002/2016EF000485

Allen, G. H., and Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, 361(6402) 585–588. doi:10.1126/science.aat0636

American Society of Civil Engineers (2017). *2017 Infrastructure Report Card: Levees*. Reston, VA. [Available at https://www.infrastructurereportcard.org/cat-item/levees/].

Archer, L., Neal, J. C., Bates, P. D., and House, J. I. (2018). Comparing TanDEM-X data with frequently used DEMs for flood inundation modeling. *Water Resources Research*, 54(12), 10205–10222. doi:10.1029/2018WR023688

Aronica, G., Bates, P. D., and Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, 16(10), 2001–2006. doi:10.1002/hyp.398

Bailly, J. S., Lagacherie, P., Millier, C., Puech, C., and Kosuth, P. (2008). Agrarian landscapes linear features detection from LiDAR: application to artificial drainage networks. *International Journal of Remote Sensing*, 29(12), 3489–3508. doi:10.1080/01431160701469057

1014 Bates, P. D., Horritt, M. S., and Fewtrell, T. J. (2010). A simple inertial formulation of the
1015 shallow water equations for efficient two-dimensional flood inundation modelling.
1016 *Journal of Hydrology*, 387(1–2), 33–45. doi:10.1016/j.jhydrol.2010.03.027

1017 Bates, P. D., Wilson, M. D., Horritt, M. S., Mason, D. C., Holden, N., and Currie, A. (2006).
1018 Reach scale floodplain inundation dynamics observed using synthetic aperture radar
1019 imagery: data analysis and modelling. *Journal of Hydrology*, 328(1–2), 306–318.
1020 doi:10.1016/j.jhydrol.2005.12.028

1021 Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M. (2016). The SWOT Mission and Its
1022 Capabilities for Land Hydrology. In: A. Cazenave, N. Champollion, J. Benveniste, and J.
1023 Chen (Eds.), *Remote Sensing and Water Resources* (pp. 117-141). Space Science Series
1024 of ISSI, vol. 55. Springer, Cham.

1025 Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. (2013). *Runoff*
1026 *Prediction in Ungauged Basins*. Cambridge, UK: Cambridge University Press.
1027 doi:10.1017/CBO9781139235761

1028 Casas, A., Riaño, D., Greenberg, J., and Ustin, S. (2012). Assessing levee stability with
1029 geometric parameters derived from airborne LiDAR. *Remote Sensing of Environment*,
1030 117, 281–288. doi:10.1016/j.rse.2011.10.003

1031 Castellarin, A., Di Baldassarre, G., Bates, P. D., and Brath, A. (2009). Optimal cross-sectional
1032 spacing in Preissmann scheme 1D hydrodynamic models. *Journal of Hydraulic*
1033 *Engineering*, 135(2), 96–105. doi:10.1061/(ASCE)0733-9429(2009)135:2(96)

1034 Castellarin, A., Di Baldassarre, G., and Brath, A. (2011a). Floodplain management strategies for
1035 flood attenuation in the river Po. *River Research and Applications*, 27(8), 1037–1047.
1036 doi:10.1002/rra.1405

1037 Castellarin, A., Domeneghetti, A., and Brath, A. (2011b). Identifying robust large-scale flood
1038 risk mitigation strategies: a quasi-2D hydraulic model as a tool for the Po river. *Physics*
1039 *and Chemistry of the Earth, Parts A/B/C*, 36(7–8), 299–308.
1040 doi:10.1016/j.pce.2011.02.008

1041 Cazorzi, F., Dalla Fontana, G., De Luca, A., Sofia, G., and Tarolli, P. (2013). Drainage network
1042 detection and assessment of network storage capacity in agrarian landscape. *Hydrological*
1043 *Processes*, 27(4), 541–553. doi:10.1002/hyp.9224

1044 Chen, B., Krajewski, W. F., Goska, R., and Young, N. (2017). Using LiDAR surveys to
1045 document floods: a case study of the 2008 Iowa flood. *Journal of Hydrology*, 553, 338–
1046 349. doi:10.1016/j.jhydrol.2017.08.009

1047 Choung, Y. (2014) Mapping levees using LiDAR data and multispectral orthoimages in the
1048 Nakdong river basins, South Korea. *Remote Sensing*, 6(9), 8696–8717.
1049 doi:10.3390/rs6098696

1050 Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., and Klijn, F. (2019). Accounting for the uncertain
1051 effects of hydraulic interactions in optimizing embankment heights: proof of principle for
1052 the Ijssel River. *Journal of Flood Risk Management*. doi:10.1111/jfr3.12532

1053 de Almeida, G. A. M., and Bates, P. (2013). Applicability of the local inertial approximation of
1054 the shallow water equations to flood modeling. *Water Resources Research*, 49(8), 4833–
1055 4844. doi:10.1002/wrcr.20366

1056 Di Baldassarre, G., Castellarin, A., and Brath, A. (2010). Analysis of the effects of levee
1057 heightening on flood propagation: example of the River Po, Italy. *Hydrological Sciences*
1058 *Journal*, 54, 1007–1017. doi:10.1623/hysj.54.6.1007

1059 Domeneghetti, A., Carisi, F., Castellarin A., and Brath, A. (2015). Evolution of flood risk over
1060 large areas: quantitative assessment for the Po river. *Journal of Hydrology*, 527, 809–823,
1061 doi:10.1016/j.jhydrol.2015.05.043

1062 Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., Feyen, L. (2016). Development
1063 and evaluation of a framework for global flood hazard mapping. *Advances in Water*
1064 *Resources*, 94, 87–102. doi:10.1016/j.advwatres.2016.05.002

1065 Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A.,
1066 Mongelli, I., Frieler, K., Betts, R. A., and Feyen, L. (2018). Increased human and
1067 economic losses from river flooding with anthropogenic warming. *Nature Climate*
1068 *Change*, 8, 781–786. doi:10.1038/s41558-018-0257-z

1069 Evans, I. S. (1979). *An integrated system of terrain analysis and slope mapping*. Final report on
1070 grant DA-ERO-591-73-G0040, University of Durham, UK.

1071 Evans, I. S. (1980). An integrated system of terrain analysis and slope mapping. *Zeitschrift für*
1072 *Geomorpologie Suppl-Bd*, 36, 274–295.

1073 Fewtrell, T. J., Bates, P. D., Horritt, M., and Hunter, N. M. (2008). Evaluating the effect of scale
1074 in flood inundation modelling in urban environments. *Hydrological Processes*, 22(26),
1075 5107–5118. doi:10.1002/hyp.7148

1076 Fewtrell, T. J., Neal, J. C., Bates, P. D., and Harrison, P. J. (2011). Geometric and structural river
1077 channel complexity and the prediction of urban inundation. *Hydrological Processes*,
1078 25(20), 3173–3186. doi:10.1002/hyp.8035

1079 Feyen, L., Dankers, R., Bódis, K., Salamon, P., and Barredo, J. I. (2012). Fluvial flood risk in
1080 Europe in present and future climates. *Climatic Change*, 112(1), 47–62.
1081 doi:10.1007/s10584-011-0339-7

1082 Fleischmann, A., Paiva, R., and Collischonn, W. (2019). Can regional to continental river
1083 hydrodynamic models be locally relevant? A cross-scale comparison. *Journal of*
1084 *Hydrology X*, 3, 100027. doi:10.1016/j.hydroa.2019.100027

1085 Frasson, R. P. d. M., Pavelsky, T. M., Fonstad, M. A., Durand, M. T., Allen, G. H., Schumann,
1086 G., Lion, C., Beighley, R. E., Yang, X. (2019). Global relationships between river width,
1087 slope, catchment area, meander wavelength, sinuosity, and discharge. *Geophysical*
1088 *Research Letters*, 46(6), 3252–3262. doi:10.1029/2019GL082027

1089 Gesch, D. B., Oimoen, M. J., and Evans, G. A. (2014). Accuracy Assessment of the US
1090 Geological Survey National Elevation Dataset, and Comparison with Other Large-Area
1091 Elevation Datasets—SRTM and ASTER. *US Geological Survey Open-File Report 2014-*
1092 *1008*. doi:10.3133/ofr20141008

1093 Gilles, D., Young, N., Schroeder, H., Piotrowski, J., and Chang, Y.-J. (2012). Inundation
1094 mapping initiatives of the Iowa Flood Center: statewide coverage and detailed urban
1095 flooding analysis. *Water*, 4(1), 85–106. doi:10.3390/w4010085

1096 Hamilton, A. S., and Moore, R. D. (2012). Quantifying uncertainty in streamflow records.
1097 *Canadian Water Resources Journal*, 37(1), 3–21. doi:10.4296/cwrj3701865

1098 Heine, R. A., and Pinter, N. (2012). Levee effects upon flood levels: an empirical assessment.
1099 *Hydrological Processes*, 26, 3225–3240. doi:10.1002/hyp.8261

1100 Hooijer, A., Klijn, F., Pedroli, G. B. M., Van Os, A. G. (2004). Towards sustainable flood risk
1101 management in the Rhine and Meuse river basins: synopsis of the findings of IRMA-
1102 SPONGE. *River Research and Applications*, 20(3), 343–357. doi:10.1002/rra.781

1103 Horna-Muñoz, D. and Constantinescu, G. (2018). A fully 3-D numerical model to predict flood
1104 wave propagation and assess efficiency of flood protection measures. *Advances in Water*
1105 *Resources*, 122, 148–165. doi:10.1016/j.advwatres.2018.10.014

1106 Horritt, M. S., Bates, P. D., Fewtrell, T. J., Mason, D. C., and Wilson, M. D. (2010). Modelling
1107 the hydraulics of the Carlisle 2005 flood event. *Proceedings of the Institution of Civil*
1108 *Engineers*, 163(WM6), 273–281. doi:10.1680/wama.2010.163.6.273

1109 Hunter, N. M., Bates, P. D., Horritt, M. S., and Wilson, M. D. (2007). Simple spatially-
1110 distributed models for predicting flood inundation: a review. *Geomorphology*, 90(3–4),
1111 208–225. doi:10.1016/j.geomorph.2006.10.021

1112 Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilla, R.,
1113 Niemeier, J., Quintero, F., Seo, B.-C., Small, S., Weber, L., and Young, N. (2017). Real-
1114 time flood forecasting and information system for the State of Iowa. *Bulletin of the*
1115 *American Meteorological Society*, 98(3), 539–554. doi:10.1175/BAMS-D-15-00243.1

1116 Krüger, T. (2010). Algorithms for detecting and extracting dikes from Digital Terrain Models.
1117 In: A. Car, G. Griesebner, J. Strobl (Eds.), *Geospatial Crossroads @ GI_Forum -10*
1118 *Proceedings of the Geoinformatics Forum Salzburg*. Berlin, Offenbach: Wichmann,
1119 S.130-139.

1120 Krüger, T., and Meinel, G. (2008). Using raster DTM for dike modelling. In: P. van Oosterom, S.
1121 Zlatanova, F. Penninga, E. M. Fendel (Eds.) *Advances in 3D Geoinformation Systems*.

Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg, doi:10.1007/978-3-540-72135-2_6.

Lamb, R., Crossley, M., and Waller, S. (2009). A fast two-dimensional floodplain inundation model. *Proceedings of the Institution of Civil Engineers – Water Management*, 162(6), 363–370. doi:10.1680/wama.2009.162.6.363

Lashermes, B., Foufoula-Georgiou, E., and Dietrich, W. E. (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters*, 34(23), L23S04. doi:10.1029/2007GL031140

Lehner, B., and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171–2186. doi:10.1002/hyp.9740

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos Transactions, American Geophysical Union*, 89(10), 93–104. doi:10.1029/2008EO100001

Masoero, A., Claps, P., Asselman, N. E. M., Mosselman, E., and Di Baldassarre, G. (2013). Reconstruction and analysis of the Po River inundation of 1951. *Hydrological Processes*, 27, 1341–1348. doi:10.1002/hyp.9558

McMillan, H., Krueger, T., and Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078–4111. doi:10.1002/hyp.9384

1142 Mignot, E., Paquier, A., and Haider, S. (2006). Modeling floods in a dense urban area using 2D
1143 shallow water equations. *Journal of Hydrology*, 327(1–2), 186–199.
1144 doi:10.1016/j.jhydrol.2005.11.026

1145 Moretti, G., and Orlandini, S. (2018). Hydrography-driven coarsening of grid digital elevation
1146 models. *Water Resources Research*, 54(5), 3654–3672. doi:10.1029/2017WR021206

1147 Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., and Horritt, M. S. (2009).
1148 Distributed whole city water level measurements from the Carlisle 2005 urban flood
1149 event and comparison with hydraulic model simulations. *Journal of Hydrology*, 368(1–
1150 4), 42–55. doi:10.1016/j.jhydrol.2009.01.026

1151 Neal, J., Dunne, T., Sampson, C., Smith, A., and Bates, P. (2018), Optimisation of the two-
1152 dimensional hydraulic model LISFLOOD-FP for CPU architecture. *Environmental*
1153 *Modelling and Software*, 107, 148–157. doi:10.1016/j.envsoft.2018.05.011

1154 Neal, J., Schumann, G., and Bates, P. (2012). A subgrid channel model for simulating river
1155 hydraulics and floodplain inundation over large and data sparse areas. *Water Resources*
1156 *Research*, 48(11), W11506. doi:10.1029/2012WR012514

1157 Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., and Matgen, P. (2007). Grasping
1158 the unavoidable subjectivity in calibration of flood inundation models: a vulnerability
1159 weighted approach. *Journal of Hydrology*, 333(2–4), 275–287.
1160 doi:10.1016/j.jhydrol.2006.08.017

1161 Pappenberger, F., Dutra, E., Wetterhall, F., and Cloke, H. L. (2012). Deriving global flood
1162 hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth*
1163 *System Sciences*, 16, 4143–4156. doi:10.5194/hess-16-4143-2012

1164 Passalacqua, P., Belmont, P., Foufoula-Georgiou, E., (2012). Automatic geomorphic feature
1165 extraction from lidar in flat and engineered landscapes. *Water Resources Research*, 48(3),
1166 W03528. doi:10.1029/2011WR010958

1167 Passalacqua, P., Tarolli, P., and Foufoula-Georgiou, E., (2010). Testing space-scale
1168 methodologies for automatic geomorphic feature extraction from lidar in a complex
1169 mountainous landscape. *Water Resources Research*, 46(11), W11535.
1170 doi:10.1029/2009WR008812

1171 Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H.,
1172 and Miller, Z. F. (2014). Assessing the potential global extent of SWOT river discharge
1173 observations. *Journal of Hydrology*, 519(B), 1516–1525.
1174 doi:10.1016/j.jhydrol.2014.08.044

1175 Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J.
1176 (2019). The spatial dependence of flood hazard and risk in the United States. *Water*
1177 *Resources Research*, 55(3), 1890–1911. doi:10.1029/2018WR024205

1178 Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E. (2015). A
1179 high-resolution global flood hazard model. *Water Resources Research*, 51(9), 7358–
1180 7381. doi:10.1002/2015WR016954

1181 Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., and Trigg, M. A. (2016). Perspectives on
1182 open access high resolution digital elevation models to produce global flood hazard
1183 layers. *Frontiers in Earth Science*, 3, 85. doi:10.3389/feart.2015.00085

1184 Sanders, B. F., and Schubert, J. E. (2019). PRIMo: parallel raster inundation model. *Advances in*
1185 *Water Resources*, 126, 79–95. doi:10.1016/j.advwatres.2019.02.007

1186 Sangireddy, H., Stark, C. P., Kladzyk, A., and Passalacqua, P. (2016). GeoNet: an open source
1187 software for the automatic and objective extraction of channel heads, channel network,
1188 and channel morphology from high resolution topography data. *Environmental Modelling*
1189 *& Software*, 83, 58–73. doi:10.1016/j.envsoft.2016.04.026

1190 Savage, J. T. S., Bates, P., Freer, J., Neal, J., and Aronica, G. (2016). When does spatial
1191 resolution become spurious in probabilistic flood inundation predictions? *Hydrological*
1192 *Processes*, 30(13), 2014–2032. doi:10.1002/hyp.10749

1193 Schmidt, A. R. (2002). *Analysis of stage-discharge relations for open-channel flow and their*
1194 *associated uncertainties*. PhD Thesis: Department of Civil and Environmental
1195 Engineering, University of Illinois at Urban-Champaign, Urbana-Champaign, IL.

1196 Schumann, G. J.-P. (2014). Fight floods on a global scale. *Nature*, 507, 169.
1197 doi:10.1038/507169e

1198 Schumann, G. J.-P., and Bates, P. D. (2018). The need for a high-accuracy, open-access global
1199 DEM. *Frontiers in Earth Science*, 6, 225. doi:10.3389/feart.2018.00225

1200 Schumann, G., Di Baldassarre, G., Alsdorf, D., and Bates, P. D. (2010). Near real-time flood
1201 wave approximation on large rivers from space: application to the River Po, Italy. *Water*
1202 *Resources Research*, 46(5), W05601. doi:10.1029/2008WR007672

1203 Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H.,
1204 and Ward, P. J. (2016). FLOPROS: an evolving global database of flood protection
1205 standards. *Natural Hazards and Earth System Sciences*, 16, 1049–1061.
1206 doi:10.5194/nhess-16-1049-2016

1207 Sofia, G., Dalla Fontana, G., and Tarolli, P. (2014). High-resolution topography and
1208 anthropogenic feature extraction: testing geomorphometric parameters in floodplains.
1209 *Hydrological Processes*, 28(4), 2046–2061. doi:10.1002/hyp.9727

1210 Sofia, G., Tarolli, P., Cazorzi, F., and Dalla Fontana, G. (2011). An objective approach for
1211 feature extraction: distribution analysis and statistical descriptors for scale choice and
1212 channel network identification. *Hydrology and Earth System Sciences*, 15, 1387–1402.
1213 doi:10.5194/hess-15-1387-2011

1214 Smith, A., Sampson, C., and Bates, P. (2015). Regional flood frequency analysis at the global
1215 scale. *Water Resources Research*, 51, 539–553. doi:10.1002/2014WR015814

1216 Steinfeld, C. M. M., Kingsford, R. T., and Laffan, S. W. (2013). Semi-automated GIS techniques
1217 for detecting floodplain earthworks. *Hydrological Processes*, 27(4), 579–591.
1218 doi:10.1002/hyp.9244

1219 Tarolli, P., Sofia, G., and Dalla Fontana, G. (2012). Geomorphic features extraction from high-
1220 resolution topography: landslide crowns and bank erosion. *Natural Hazards*, 61(1), 65–
1221 83. doi:10.1007/s11069-010-9695-2

1222 Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., Yamazaki, D.,
1223 Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P. J., Winsemius, H. C., Salamon, P.,
1224 Dottori, F., Rudari, R., Kappes, M. S., Simpson, A. L., Hadzilacos, G., and Fewtrell, T. J.
1225 (2016). The credibility challenge for global fluvial flood risk analysis, *Environmental*
1226 *Research Letters*, 11, 094014. doi:10.1088/1748-9326/11/9/094014

1227 van der Made, J. E. (1982). Determination of the accuracy of water level observations.
1228 *Proceedings of the Exeter Symposium, IAHS Publications 134*, 172–184.

1229 Villarini, G., Serinaldi, F., Smith, J. A., and Krajewski, W. F. (2009a). On the stationarity of
1230 annual flood peaks in the continental United States during the 20th century. *Water*
1231 *Resources Research*, 45(8). doi:10.1029/2008WR007645.

1232 Ward, P. J., Jongman, B., Aerts, J. C. J. H., Bates, P. D., Botzen, W. J. W., Loaiza, A. D.,
1233 Hallegatte, S., Kind, J. M., Kwadijk, J., Scussolini, P., and Winsemius, H. C. (2017). A
1234 global framework for future costs and benefits of river-flood protection in urban areas.
1235 *Nature Climate Change*, 7, 642–646. doi:10.1038/nclimate3350

1236 Werner, M. G. F., Hunter, N. M., and Bates, P. D. (2005). Identifiability of distributed floodplain
1237 roughness values in flood extent estimation. *Journal of Hydrology*, 314(1–4), 139–157.
1238 doi:10.1016/j.jhydrol.2005.03.012

1239 Wilson, M., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., Frappart, F., and
1240 Famiglietti, J. (2007). Modeling large-scale inundation of Amazonian seasonally flooded
1241 wetlands. *Geophysical Research Letters*, 34(15), L15404. doi:10.1029/2007GL030156

1242 Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. J., and Erickson, T. A.
1243 (2017). Validation of a 30 m resolution flood hazard model of the conterminous United
1244 States. *Water Resources Research*, 53(9), 7968–7986. doi:10.1002/2017WR020917

1245 Wing, O., Bates, P., Sampson, C., Smith, A., Fargione, J., and Johnson, K. (2018). Millions more
1246 Americans face flood risks than previously thought. *Eos*, 99. doi:10.1029/2018EO106321

1247 Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., and Bouwman, A. (2013). A
1248 framework for global river flood risk assessments. *Hydrology and Earth System Sciences*,
1249 17, 1871–1892. doi:10.5194/hess-17-1871-2013

1250 Winsemius, H. C., Ward, P. J., Gayton, I., Veldhuis, M.-C., Meiher, D. H., and Iliffe, M. (2019).
1251 Commentary: the need for high-accuracy, open-access global DEM. *Frontiers in Earth*
1252 *Science*, 7, 33. doi:10.3389/feart.2019.00033

1253 Wood, J. D. (1996). *The geomorphological characteristics of digital elevation models.* PhD
1254 Thesis: Department of Geography, University of Leicester.

1255 Wood, M., de Jong, S. M., and Straatsma, M. W. (2018). Locating flood embankments using
1256 SAR time series: a proof of concept. *International Journal of Applied Earth Observation*
1257 *and Geoinformation*, 70, 72–83. doi:10.1016/j.jag.2018.04.003

1258 Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson,
1259 C. C., Kanae, S., and Bates, P. D. (2017). A high-accuracy map of global terrain
1260 elevations. *Geophysical Research Letters*, 44(11), 5844–5853.
1261 doi:10.1002/2017GL072874

1262 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M. (2019).
1263 MERIT Hydro: a high-resolution global hydrography map based on latest topography
1264 dataset. *Water Resources Research*, 55. doi:10.1029/2019WR024873

1265 Yamazaki, D., Kanae, S., Kim, H., and Oki, T. (2011). A physically based description of
1266 floodplain inundation dynamics in a global river routing model. *Water Resources*
1267 *Research*, 47(4), W04501. doi:10.1029/2010WR009726

1268 Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., and Alsdorf, D. E. (2012).
1269 Estimating river bathymetry from data assimilation of synthetic SWOT measurements.
1270 *Journal of Hydrology*, 464–465, 363–375. doi:10.1016/j.jhydrol.2012.07.028

**Figure 1 (.eps).**

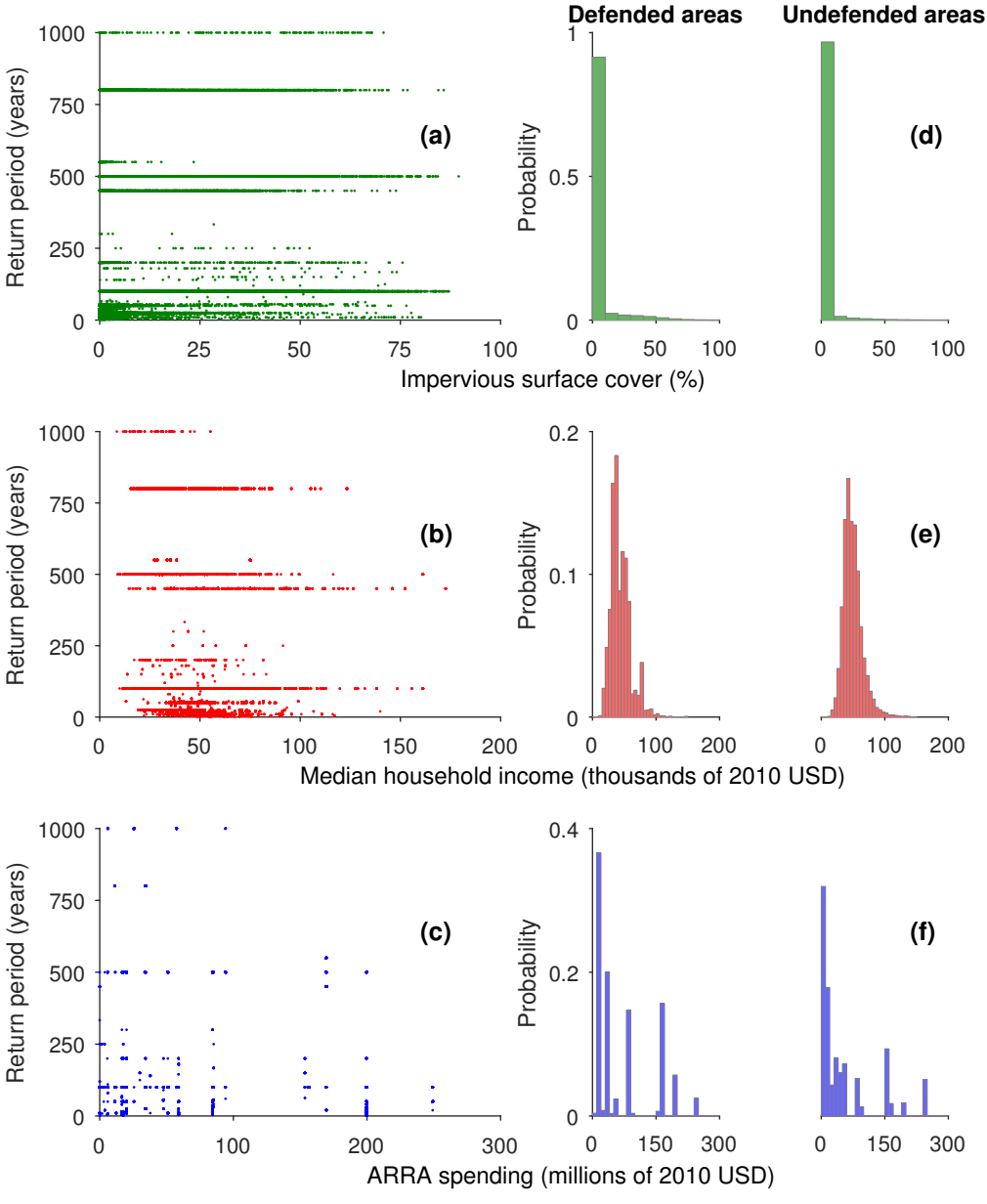

Defended areas
Undefended areas
(a)
(b)
(c)
(d)
(e)
(f)
Return period (years)
Probability
Impervious surface cover (%)
Median household income (thousands of 2010 USD)
ARRA spending (millions of 2010 USD)

**Figure 2 (.eps).**

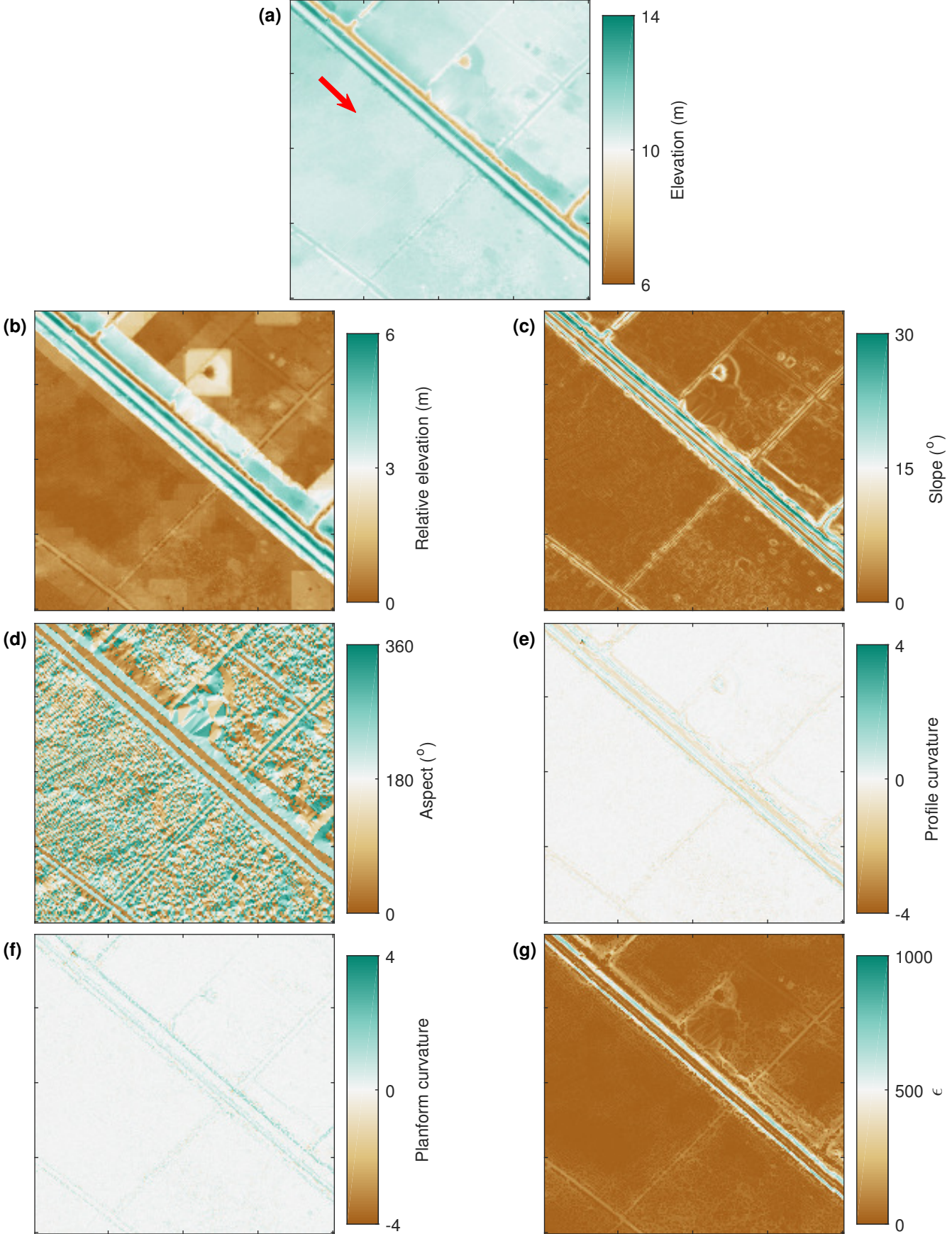

(a)
14
10
6
Elevation (m)
(b)
6
3
0
Relative elevation (m)
(c)
30
15
0
Slope (°)
(d)
360
180
0
Aspect (°)
(e)
4
0
-4
Profile curvature
(f)
4
0
-4
Planform curvature
(g)
1000
500
0
ϵ

**Figure 3 (.eps).**

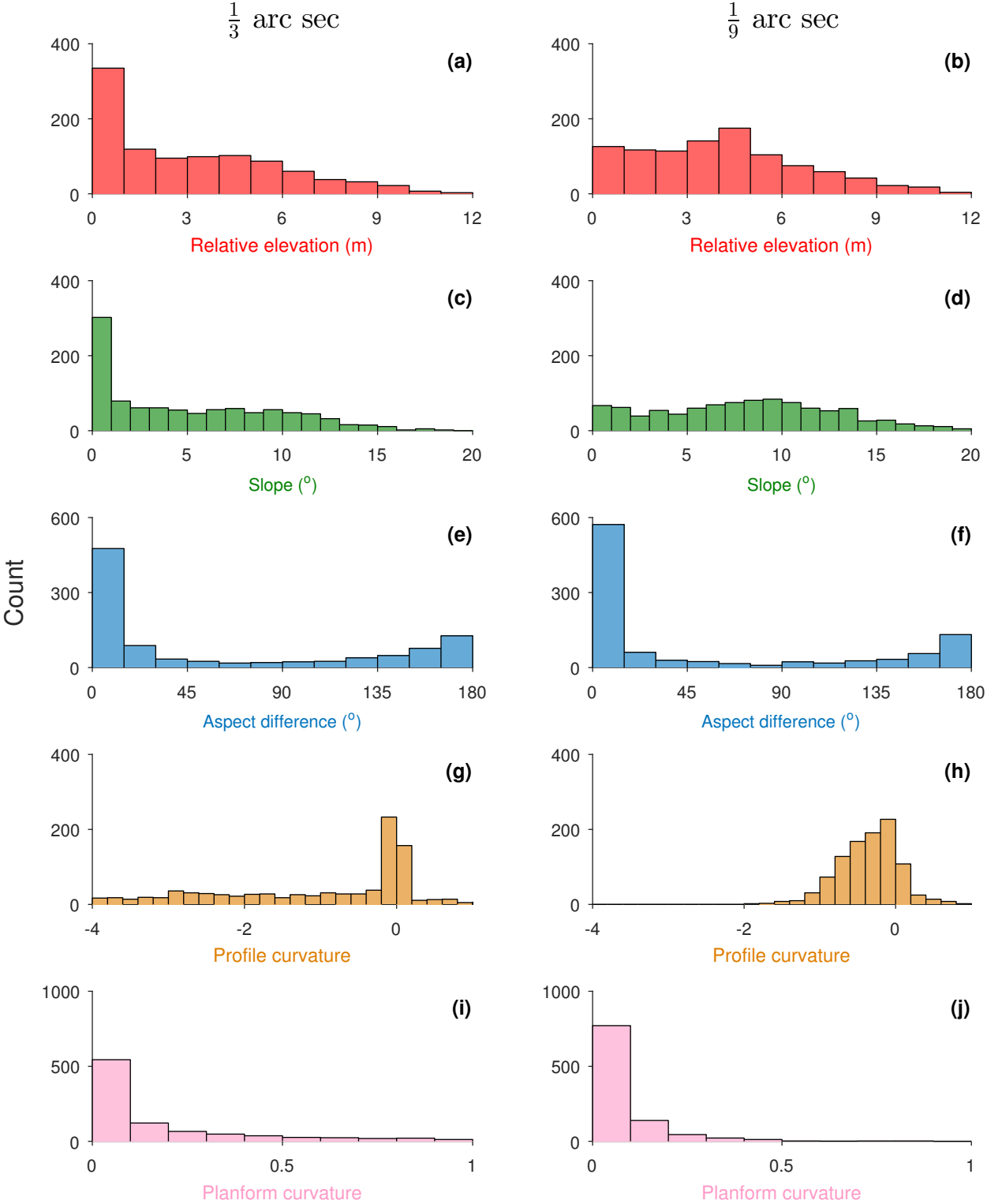

1/3 arc sec
1/9 arc sec
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
Count
Relative elevation (m)
Slope (°)
Aspect difference (°)
Profile curvature
Planform curvature
400
200
0
600
300
1000
500
0
3
6
9
12
5
10
15
20
45
90
135
180
-4
-2
0.5
1

**Figure 4 (.eps).**

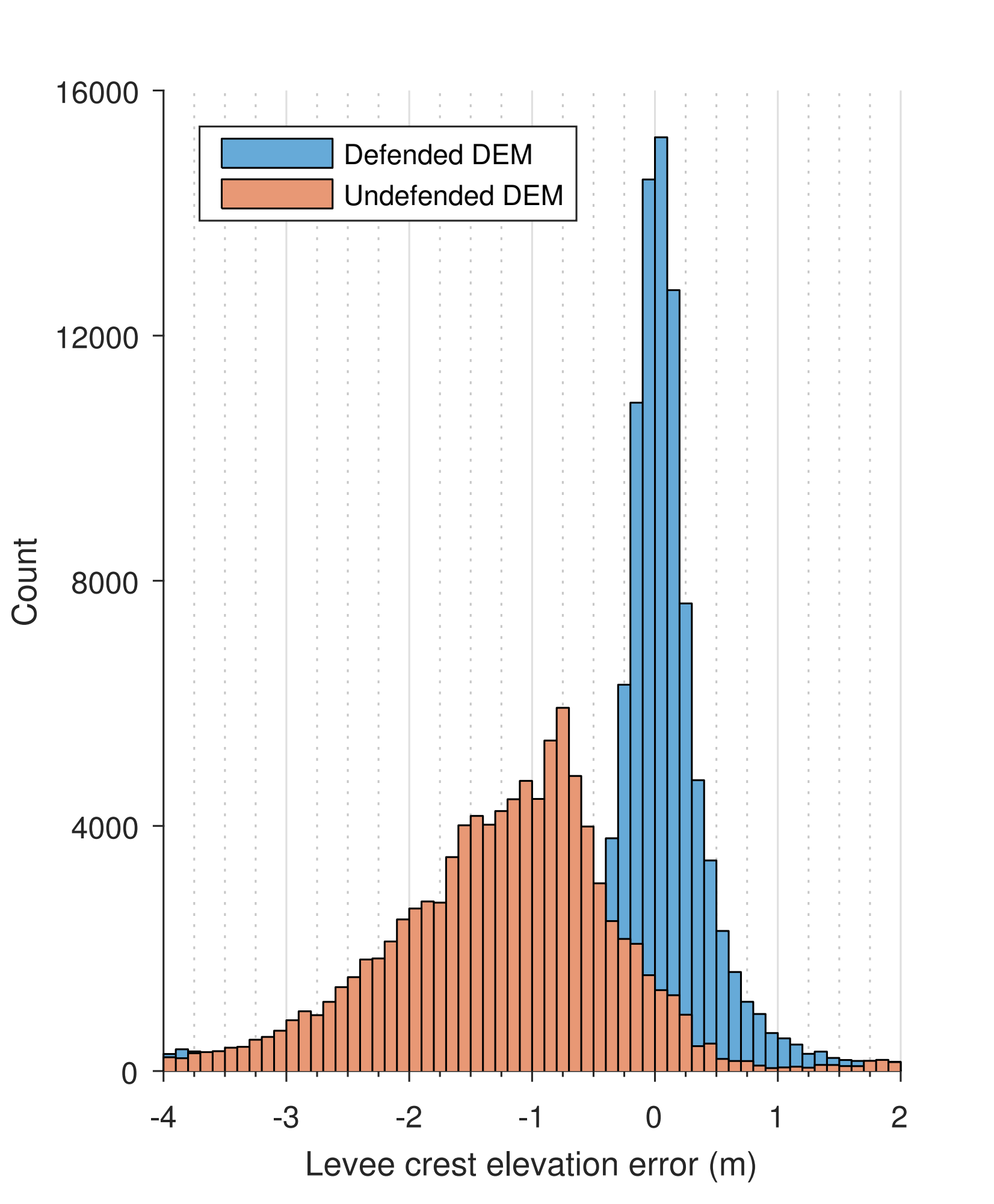

Defended DEM
Undefended DEM
16000
12000
8000
4000
0
Count
-4
-3
-2
-1
0
1
2
Levee crest elevation error (m)

**Figure 5 (.eps).**

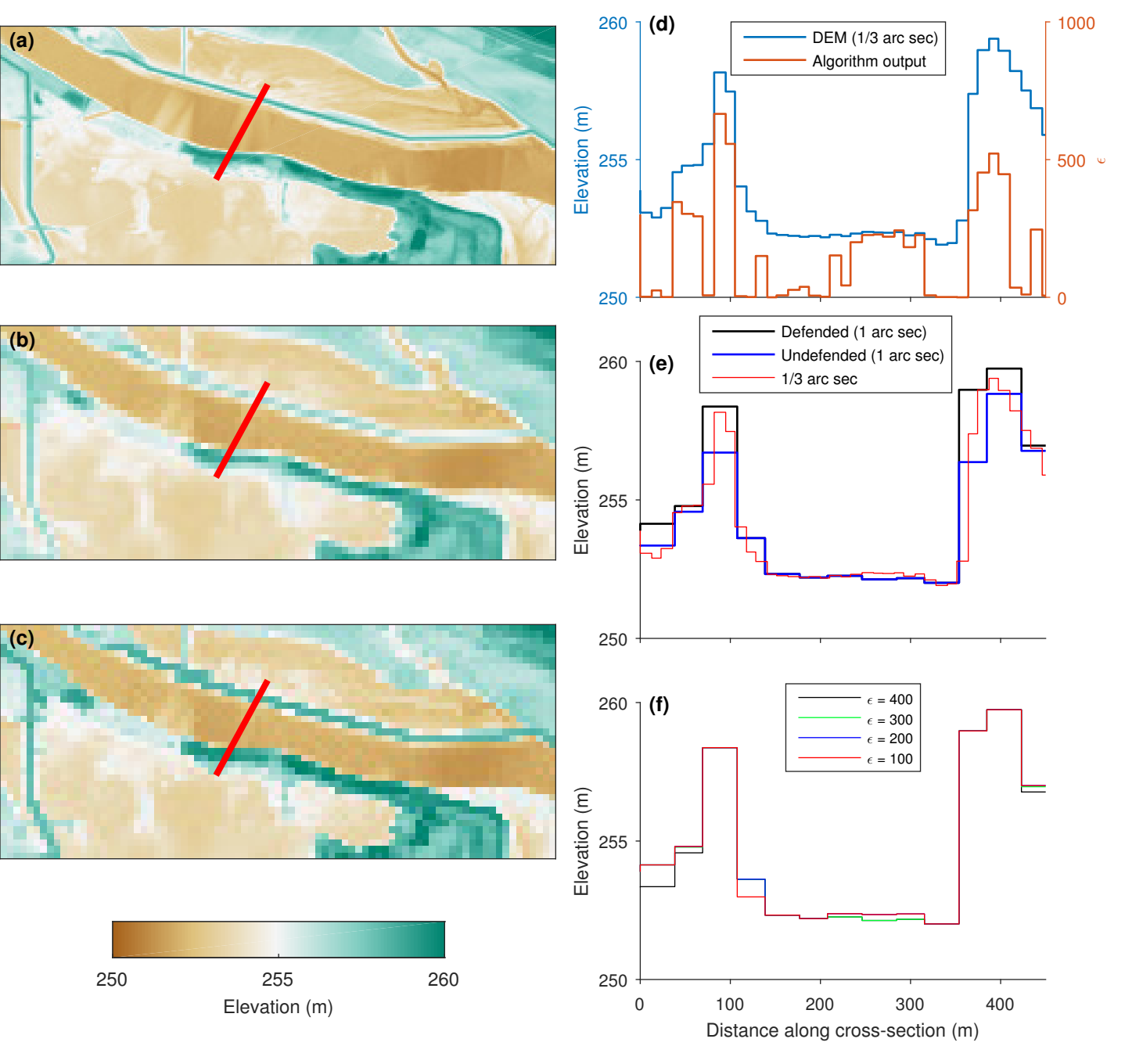

(a)
(b)
(c)
(d)
(e)
(f)
250
255
260
Elevation (m)
DEM (1/3 arc sec)
Algorithm output
1000
500
0
ϵ
Defended (1 arc sec)
Undefended (1 arc sec)
1/3 arc sec
ϵ = 400
ϵ = 300
ϵ = 200
ϵ = 100
0
100
200
300
400
Distance along cross-section (m)

**Figure 6 (.tif).**

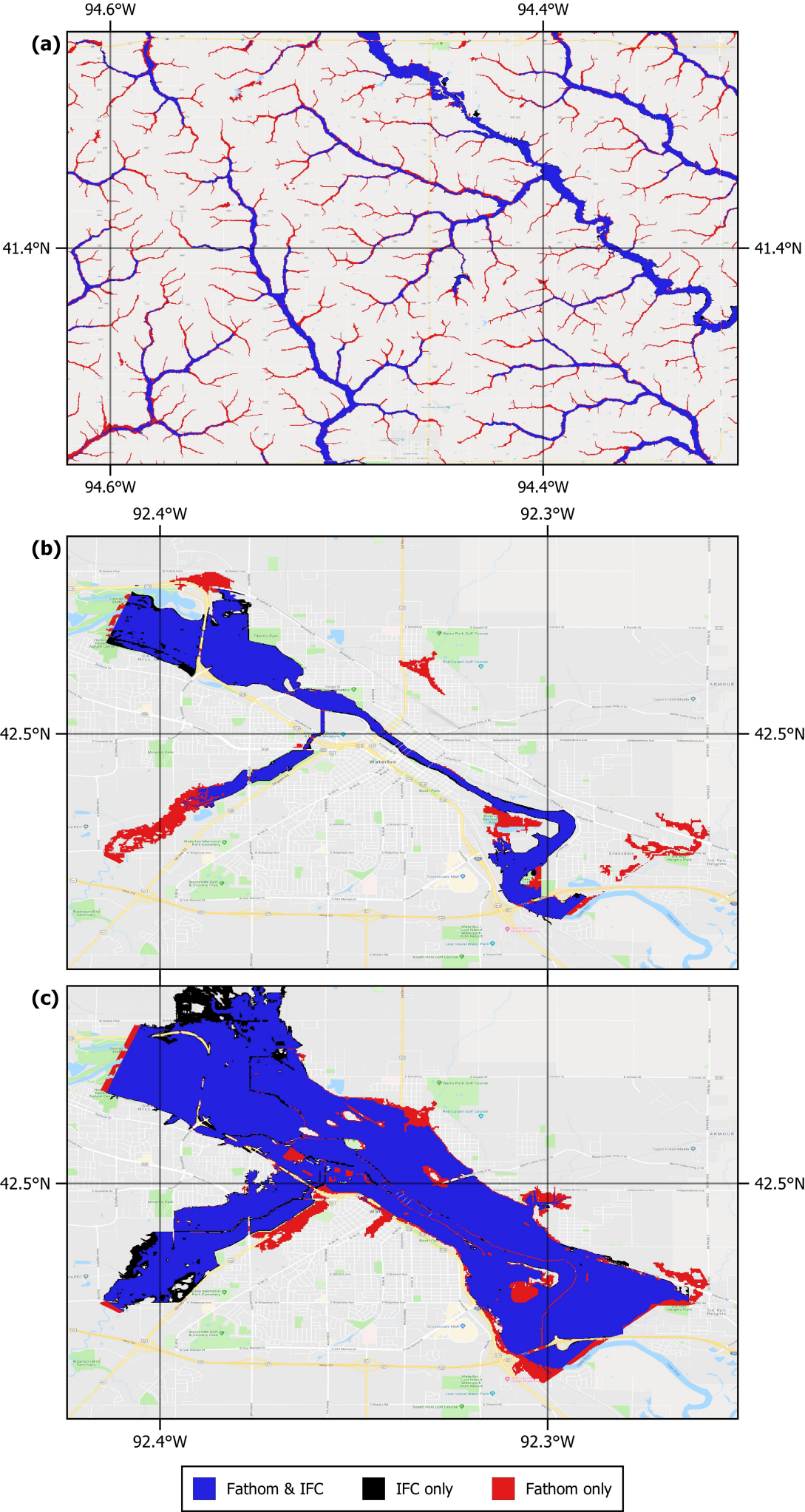

(a)
94.6°W
94.4°W
41.4°N
41.4°N
94.6°W
94.4°W
(b)
92.4°W
92.3°W
42.5°N
42.5°N
(c)
42.5°N
42.5°N
92.4°W
92.3°W
Fathom & IFC
IFC only
Fathom only

**Figure 7 (.eps).**

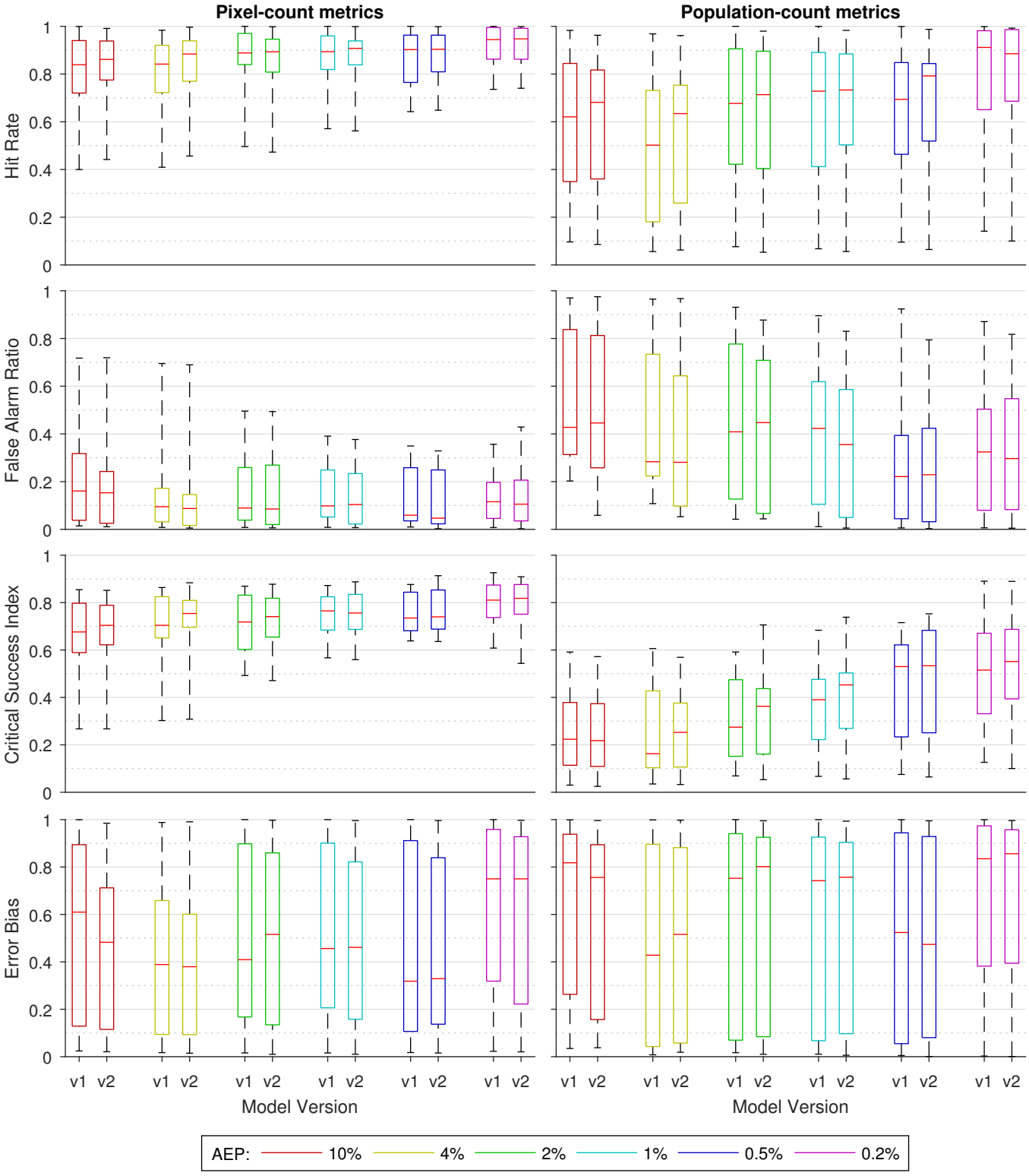

Pixel-count metrics
Population-count metrics
Hit Rate
False Alarm Ratio
Critical Success Index
Error Bias
0
0.2
0.4
0.6
0.8
1
v1 v2
Model Version
AEP: 10% 4% 2% 1% 0.5% 0.2%

**Figure 8 (.tif).**

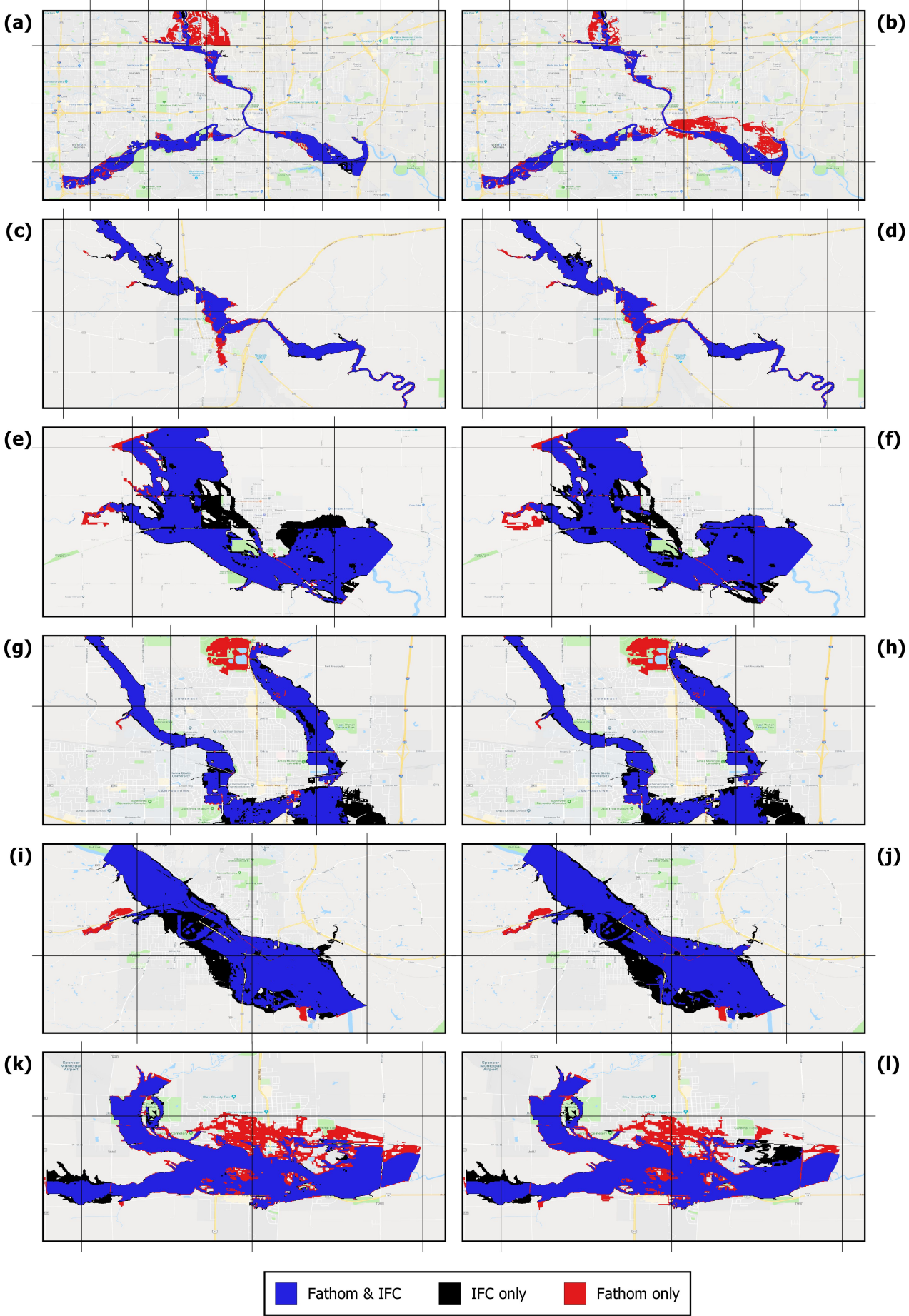

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Fathom & IFC
IFC only
Fathom only

**Figure 9 (.tif).**

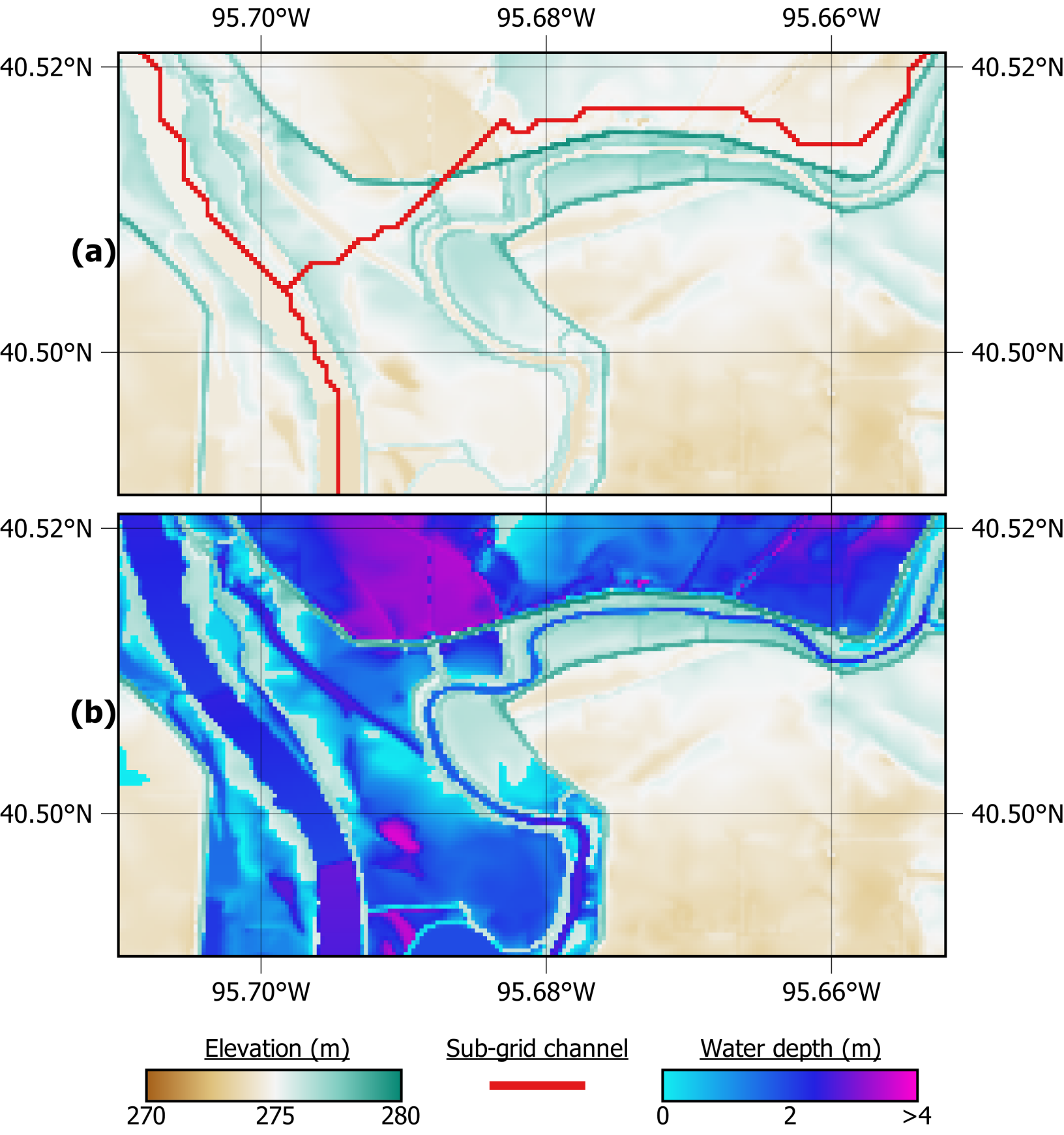

95.70°W
95.68°W
95.66°W
40.52°N
40.50°N
(a)
(b)
Elevation (m)
270
275
280
Sub-grid channel
Water depth (m)
0
2
>4

**Figure 10 (.tif).**

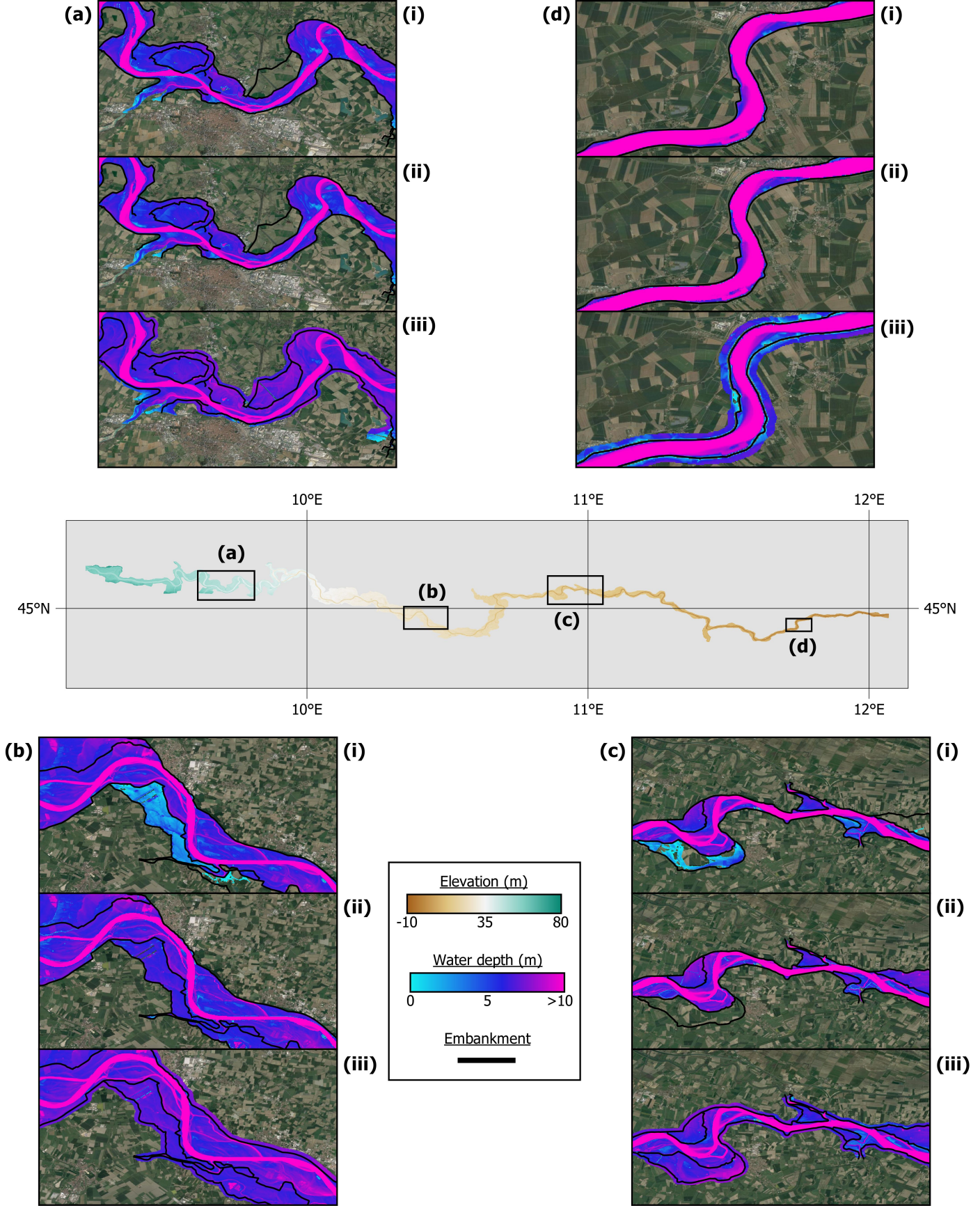

(a)
(i)
(ii)
(iii)
(d)
(i)
(ii)
(iii)
10°E
11°E
12°E
(a)
(b)
(c)
(d)
45°N
45°N
10°E
11°E
12°E
(b)
(i)
(ii)
(iii)
Elevation (m)
-10
35
80
Water depth (m)
0
5
>10
Embankment
(c)
(i)
(ii)
(iii)

**Figure 11 (.tif).**

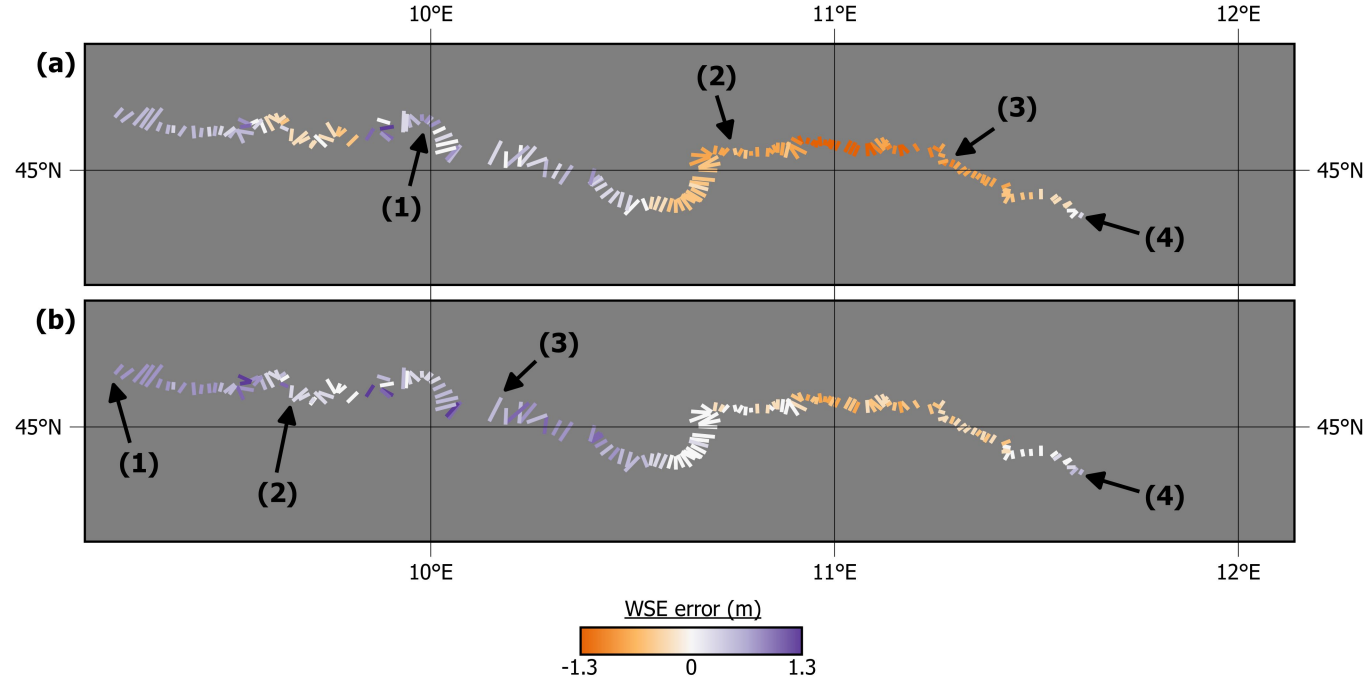

(a)
(b)
10°E
11°E
12°E
45°N
(1)
(2)
(3)
(4)
WSE error (m)
-1.3
0
1.3

**Figure 12 (.tif).**

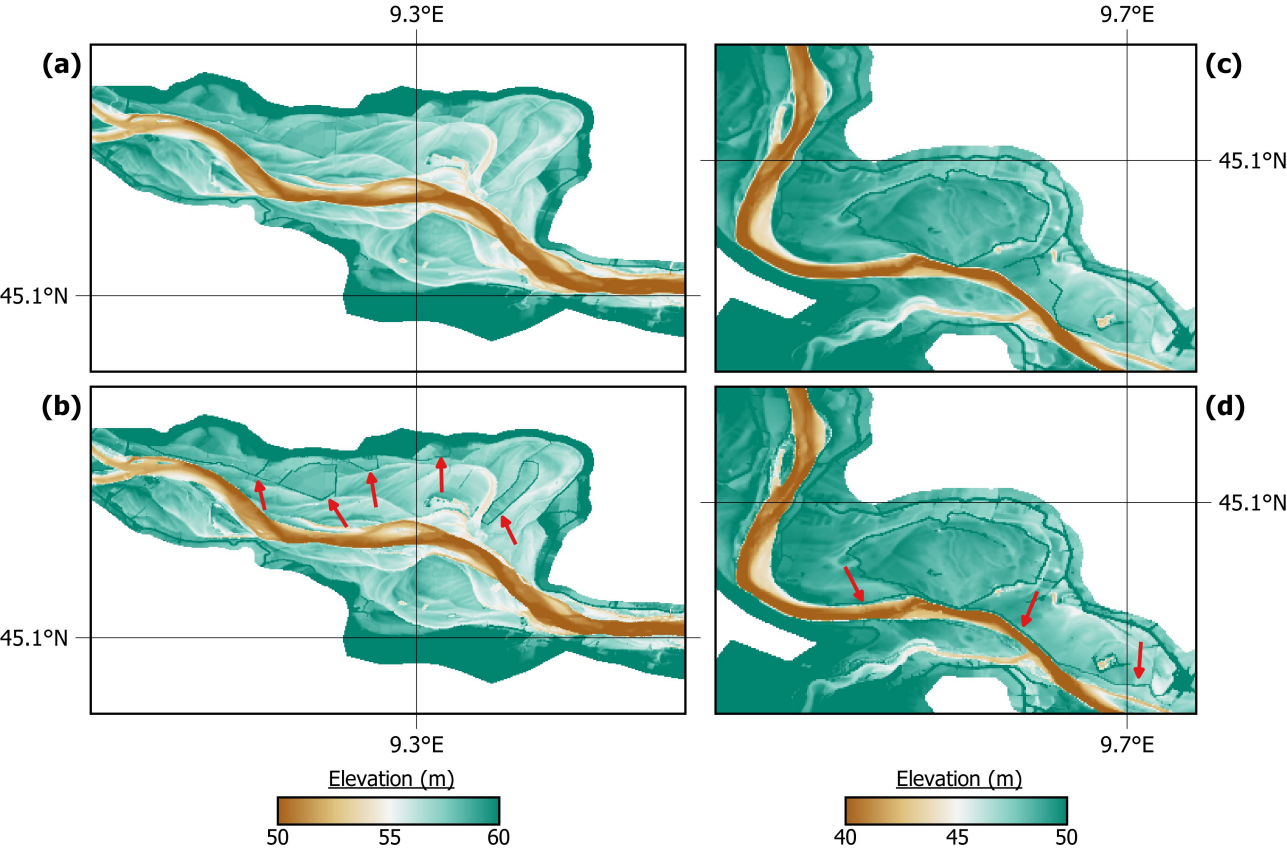

9.3°E
9.7°E
(a)
(b)
(c)
(d)
45.1°N
45.1°N
45.1°N
45.1°N
9.3°E
9.7°E
Elevation (m)
50
55
60
Elevation (m)
40
45
50

**Figure 13 (.tif).**

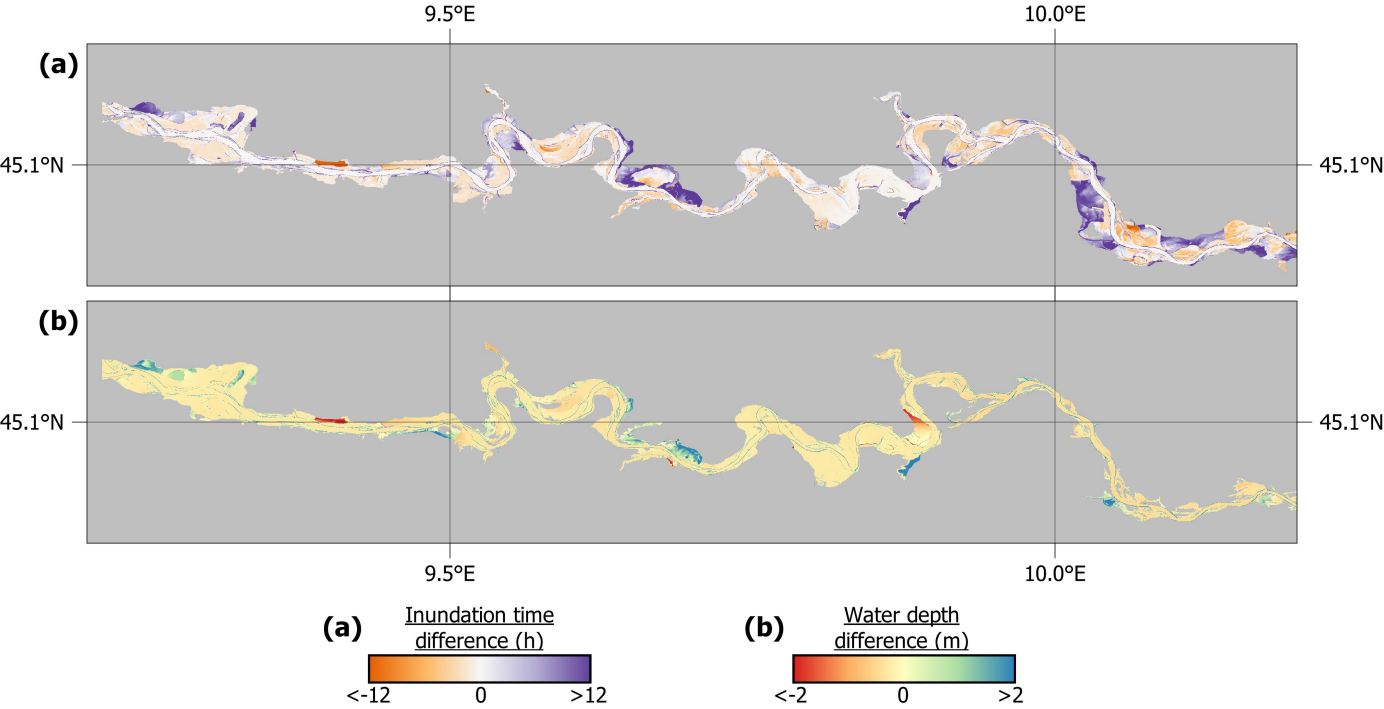
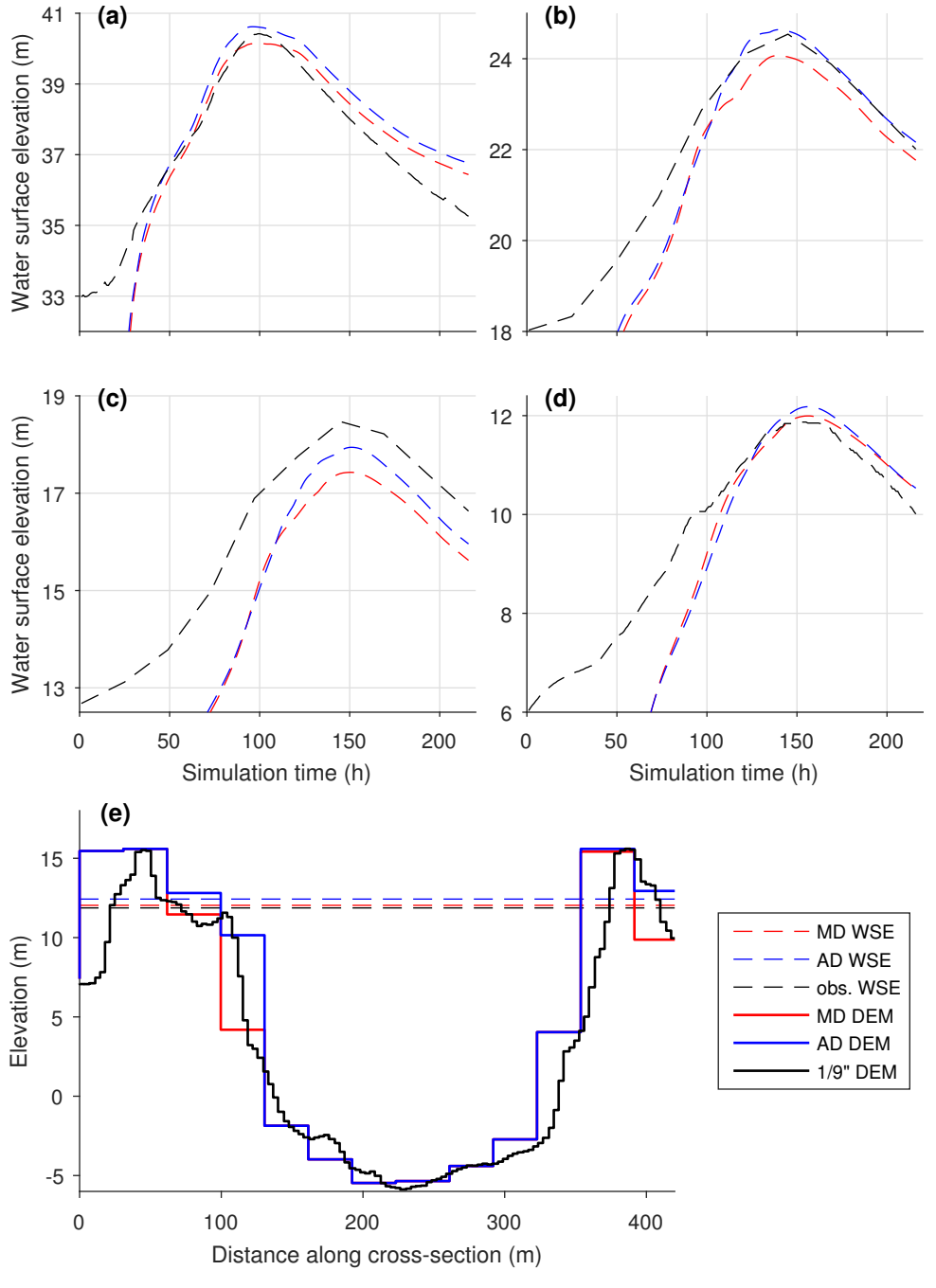

9.5°E
10.0°E
(a)
45.1°N
45.1°N
(b)
45.1°N
45.1°N
9.5°E
10.0°E
(a)
Inundation time difference (h)
<-12
0
>12
(b)
Water depth difference (m)
<-2
0
>2

**Figure 14 (.eps).**


(a)
(b)
(c)
(d)
(e)
Water surface elevation (m)
Simulation time (h)
Elevation (m)
Distance along cross-section (m)
MD WSE
AD WSE
obs. WSE
MD DEM
AD DEM
1/9" DEM